IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*,

Plaintiffs,

v.

RICK PERRY, *et al.*,

Defendants.

Civil Actions No. 2:13-cv-193 (NGR)

DECLARATION OF STEPHEN D. ANSOLABEHERE

Pursuant to 28 U.S.C. § 1746, I, Stephen D. Ansolabehere, make the following declaration:

Expert Report

Stephen Ansolabehere

JULY 1, 2014

I. Statement of Inquiry

1. I have been asked to identify which registered voters in the State of Texas have an acceptable photographic identification required by Texas Senate Bill 14 (2011) (SB 14) and which do not.

2. I was also asked to determine whether there is a disparity between the percent of Anglo registered voters and Black or Hispanic registered voters who possess acceptable SB 14 photo ID.

3. In order to determine the number of Texas registered voters without acceptable SB 14 photo ID, and to determine any racial disparities in rates of acceptable SB 14 ID between Blacks, Hispanics, and Anglos, I was retained to match records in the Texas Election Administration Management system (TEAM) voter registration database to relevant State of Texas and Federal identification databases.

4. I analyzed rates of ID possession for registered voters by race in two different ways. This is necessary because the TEAM voter registration database does not contain self-reported information on the race of each registered voter. First, I conducted an ecological regression analysis that relates rates of possession of acceptable SB 14 ID among registered voters to the racial composition of Census areas. Second, I matched individual voter records to data from a firm, Catalist, LLC, that uses a statistical model to predict the race of individual registered voters.

5. Separate from the determination of who has acceptable SB 14 photo ID, I have also been asked to examine historical rates of registration and voting among racial groups. I did so in three ways: (i) using data from the Bureau of Census's Current Population Survey (CPS); (ii) by ecological regression analysis relating registration and voting rates to the racial composition of Census areas; and (iii) by analysis of TEAM voter history data combined with the Catalist classification of the race of voters.

6. In brief, I conclude that:

(a) There are approximately 1.2 million voters in Texas who do not possess acceptable SB 14 photo identification, representing 9.1 percent of registered voters. Moreover, approximately 1.1 million voters in Texas neither possess acceptable SB 14 photo ID nor qualify under SB 14 to apply for a disability-based exemption from showing ID at the polls.

(b) There is a racial disparity in the rates of possession of acceptable SB 14 photo ID such that Black[1] registered voters are approximately twice as likely as Anglo registered voters to lack acceptable photo ID and that Hispanic registered voters are approximately fifty percent more likely than Anglo registered voters to lack acceptable SB 14 photo ID.

(c) In recent elections examined, rates of both voter registration and voter turnout are lower for Blacks and Hispanics in Texas than Anglos.

[1] Throughout, Black refers to individuals who are Black non-Hispanic.

II. Background and Qualifications

7. I am a professor of Government in the Department of Government at Harvard University in Cambridge, MA. Formerly, I was an Assistant Professor at the University of California, Los Angeles, and I was Professor of Political Science at the Massachusetts Institute of Technology, where I held the Elting R. Morison Chair and served as Associate Head of the Department of Political Science. I directed the Caltech/MIT Voting Technology Project from its inception in 2000 through 2004, am the Principal Investigator of the Cooperative Congressional Election Study, a survey research consortium of over 250 faculty and student researchers at more than 50 universities, and serve on the Board of Overseers of the American National Election Study. I am a consultant to CBS News' Election Night Decision Desk. I am a member of the American Academy of Arts and Sciences (inducted in 2007).

8. I have worked as a consultant to the Brennan Center in the case of *McConnell v. FEC*, 540 U.S. 93 (2003). I have testified before the U.S. Senate Committee on Rules, the U.S. Senate Committee on Commerce, the U.S. House Committee on Science, Space, and Technology, the U.S. House Committee on House Administration, and the Congressional Black Caucus on matters of election administration in the United States. I filed an amicus brief with Professors Nathaniel Persily and Charles Stewart on behalf of neither party to the U.S. Supreme Court in the case of *Northwest Austin Municipal Utility District Number One v. Holder*, 557 U.S. 193 (2009). I am consultant for the Rodriguez plaintiffs in *Perez v. Perry*, currently before the U. S.

District Court in the Western District of Texas (No. 5:11-cv-00360), and the Gonzales intervenors in *State of Texas v. United States* before the U.S. District Court in the District of Columbia (No. 1:11-cv-01303); I consulted for the Department of Justice in *State of Texas v. Holder*, before the U.S. District Court in the District of Columbia (No. 1:12-cv-00128); I consulted for the Guy plaintiffs in *Guy v. Miller* in U.S. District Court for Nevada (No. 11-OC-00042-1B); I consulted for the Florida Democratic Party in *In re Senate Joint Resolution of Legislative Apportionment* in the Florida Supreme Court (Nos. 2012-CA-412, 2012-CA-490); I am consultant for the Romo plaintiffs in *Romo v. Detzner* in the Circuit Court of the Second Judicial Circuit in Florida (No. 2012 CA 412); I am consultant for the San Antonio Water District intervenor in *LULAC v. Edwards Aquifer Authority* in the U.S. District Court for the Western District of Texas, San Antonio Division (No. 5:12cv620-OLG,); I am consultant for the Harris plaintiffs in *Harris v. McCrory* in the U. S. District Court for the Middle District of North Carolina (No. 1:2013cv00949).

9. My areas of expertise include American government, with particular expertise in electoral politics, representation, and public opinion, as well as statistical methods in social sciences. I have authored numerous scholarly works on voting behavior and elections, the application of statistical methods in social sciences, legislative politics and representation, and distributive politics. This scholarship includes articles in such academic journals as the Journal of the Royal Statistical Society, the American Political Science Review, the American Economic Review, the American Journal of Political Science, Legislative Studies Quarterly, the Quarterly Journal of

Political Science, Electoral Studies, and Political Analysis. I have published articles on issues of election law in the Harvard Law Review, Texas Law Review, Columbia Law Review, New York University Annual Survey of Law, and the Election Law Journal, for which I am a member of the editorial board. I have coauthored three scholarly books on electoral politics in the United States, The End of Inequality: Baker v. Carr and the Transformation of American Politics, Going Negative: How Political Advertising Shrinks and Polarizes the Electorate, and The Media Game: American Politics in the Media Age. I am coauthor with Ted Lowi, Ben Ginsberg, and Ken Shepsle of American Government: Power and Purpose. My curriculum vita with publications list is attached to this report.

10. As the Principal Investigator of the Cooperative Congressional Election Study and the Harvard Election Data Archive, I have extensive experience with database management, record linkage and database matching, data validation, and integration of Census and electoral data. I have published articles in refereed journals on matching survey data to voter files, on validation of voting records, and on statistical techniques for analyzing aggregate election and population data.

11. I have been hired by the Department of Justice in this case. I am retained for a rate of $400 per hour, which is my standard consulting rate.

III. Summary of Analysis and Findings

12. I have been asked to determine the number of Texas registered voters who lack acceptable SB 14 photo ID and to determine whether there are disparities between the percentages of Anglo registered voters and of Black or Hispanic registered voters who possess such identification and those who neither possess such identification nor qualify for SB 14's disability exemption from providing identification at the polls.

13. In order to do this, I conducted database matching and record linkage of the Texas Election Administration Management system (TEAM) database to relevant State of Texas and Federal databases. SB 14 specifies 7 categories of state and Federal identification that may be used when voting.[2] In addition, SB 14 allows persons with certain Federally-documented disabilities to apply for an exemption from showing identification at the polls.[3] The TEAM database records which voters have applied for and received this disability exemption. Using standard methods for

[2] Specifically, SB 14 requires that all in-person voters present one of the following in order to cast a regular ballot: (1) a Texas Driver License (DL); (2) a Texas Personal Identification Card (PID); (3) a Texas concealed handgun license (CHL); (4) a Texas Election Identification Certificate (EIC); (5) a U.S. military identification card with photo; (6) a U.S. citizenship certification (defined to include certificates of naturalization) with photo; or (7) a U.S. passport. SB 14 requires that IDs with an expiration date be current or expired for less than 60 days.

[3] To apply for the disability exemption, voters must submit written documentation stating that they do not have an acceptable SB 14 photo ID and showing either that they have been determined to be disabled by the Social Security Administration, or that they have a disability rating from the U.S. Department of Veteran Affairs of at least 50%. As of January 15, 2014, only 18 Texas voters had been approved for a disability exemption from showing acceptable SB 14 ID.

linking databases, I matched records in the TEAM voter registration database to records of holders of Texas-issued forms of SB 14 ID contained in the Texas Department of Public Safety (DPS) databases. Also, I developed protocols for matching the TEAM voter registration database to databases of the Federal agencies that issue allowable forms of SB 14 identification, as well as the databases of the Federal agencies that make relevant disability determinations.[4]

14. The matching algorithm produces a NO MATCH list, which consists of all records on TEAM for which no matching record could be found in any identification database and which are not recorded in TEAM as having received a disability exemption. Each record on this list is treated as an individual registered voter who lacks acceptable SB 14 photo ID. The algorithm also produces a MATCH list, which consists of all records on TEAM for which a matching record could be found on an identification database.

15. The matching algorithm also produces a NO MATCH/NOT EXEMPTION ELIGIBLE list, which consists of all records in TEAM for which no matching record could be found in any state or Federal identification database, which are not recorded in TEAM as having received a disability exemption, and for which no matching record in the relevant Federal disability databases could be found. Each record on this list

[4] The Federal agencies that issue allowable SB 14 photo ID and which matched their databases to TEAM are: the Department of Defense (DOD), Department of State (DOS), Department of Veterans Affairs (VA), and the United States Citizenship and Immigration Services (USCIS). The Federal agencies that make the relevant disability determinations are the Social Security Administration (SSA) and the VA.

is treated as an individual registered voter who lacks acceptable SB 14 photo ID and does not qualify for SB 14's disability exemption.

16. The TEAM data extracted on January 15, 2014, contain 13,564,416 records. Of these records, 1,232,242 were not matched to any record in a State of Texas or Federal identification database or had not already received the disability exemption.[5] These records make up the NO MATCH list for my analysis. And, 1,117,496 in TEAM had not received a disability exemption, were not matched to a state or Federal identification, and were not matched to a Federal disability database. These 1,117,496 records make up the NO MATCH/NOT EXEMPTION ELIGIBLE list.

17. I performed two analyses of racial disparities in the incidence of inclusion on the NO MATCH and NO MATCH/NOT EXEMPTION ELIGIBLE lists. First, the likelihood that a registered voter is on the NO MATCH list is correlated with racial data from the Census Bureau, using ecological regression. Ecological regression is widely used to measure the voting patterns of racial groups in Voting Rights Act cases. Second, the likelihood that an individual record in TEAM is on either the NO MATCH list or the NO MATCH/NOT EXEMPTION ELIGIBLE list is related to the individual's race, using the classification of race of registered voters provided by Catalist, LLC, a data utility company that provides election data.

[5] Only 18 registered voters in TEAM were marked as having already received the disability exemption.

18. These two forms of analysis for the NO MATCH list yield very similar results and show that there are statistically significant[6] disparities between the rate at which Anglos in TEAM are matched to identification databases and disability data records and the rates at which Blacks and Hispanics are matched. Analysis of the NO MATCH/NOT EXEMPTION ELIGIBLE list shows that the racial disparities found as to the NO MATCH list are not alleviated upon considering eligibility to apply for the disability exemption.

19. Ecological regression based on the NO MATCH list estimates that 5.3% percent of registered Anglos do not match to a corresponding record in a state or federal identification database. By comparison, an estimated 13.2 percent of registered Blacks and 8.8 percent of registered Hispanics do not match. (See Table VI.1.)

20. Analysis of the NO MATCH list using the Catalist racial classification estimates shows that 7.4 percent of registered Anglos do not match to a corresponding record in a state or federal identification database. By comparison, an estimated 15.1 percent of registered Blacks and 11.3 percent of registered Hispanics do not match. (See Table VI.2.)

[6] Throughout I report as "statistically significant differences" results for which a statistical test yields a probability of less than 1 percent that the observed difference arose by chance (i.e. that the true difference is 0). This corresponds to the observation that the hypothesized value of 0 is outside of the 99 percent confidence interval for the observed difference.

21. Analysis of the NO MATCH/NOT EXEMPTION ELIGIBLE list using the Catalist racial classification estimates shows that 6.8 percent of registered Anglos do not match to a corresponding record in a state or federal identification database, nor in a Federal disability database. By comparison, an estimated 13.2 percent of registered Blacks and 10.4 percent of registered Hispanics do not match in the relevant databases. (See Table VI.3.)

22. Further analyses validate these findings and test the robustness of the results to alternative specifications of the pool of registered voters and the classification of individuals' races. The pattern of results remains consistent across these alternate analyses, with each analysis showing statistically significant racial disparities in rates of matching of records in TEAM to records in acceptable SB 14 photo ID databases between Anglo registered voters and Black and Hispanic registered voters. (See Section VII.)

23. I also examine voting and registration patterns of racial groups in the State of Texas from 2006 to 2012 to determine whether racial differences in participation exist currently or historically. I examine three data sets: TEAM data merged with Catalist racial classifications, the Current Population Survey (CPS) reports on voting and registration rates, and aggregate data on registration and voting and racial composition of populations. All three data sets show that Anglos in the State of Texas register and vote at higher rates than minorities. (See Section VIII.)

IV. Data and Sources

A. Terminology

24. TEAM Database. The official list of registered voters in the State of Texas, maintained by the Texas Secretary of State. The Texas Election Administration Management (TEAM) System includes registered voter information such as name, address, date of birth, gender, and past elections in which the person voted.

Identification Database. A state or Federal list of individuals with a given form of identification, such as a Driver License or Passport, which includes names, addresses, and other information about the individuals.

Disability Database. A Federal list of individuals with a given disability status, which includes names, addresses, and other information about the individuals.

Record. A row in a database containing the information in that database for a specific person. Also called a case.

Field. A column in a database corresponding to information about each of the records in the database, such as first names or dates of birth.

Identifier. A field or constructed combination of fields for a particular record in a database that can be used to identify another record likely to be the same person in a different database.

Unique Identifier. An identifier that is unique to a given individual.

Record linkage. A process whereby a given record in one database is matched to one or more records in other databases using identifiers for individual records. Also called matching.

Match. A record in a database found to have at least one matching identifier in a separate database.

No Match. A record for which no match is found in other databases.

Sweep. A search conducted of all records in one database using a specified identifier for matching records in another database.

MATCH List. Records for all currently registered voters in TEAM for which at least one matching record is found in an identification database on at least one sweep.

NO MATCH List. Records for all currently registered voters in TEAM which do not list the voter as having already applied for and received the disability exemption and for which no matching record is found in any identification database on any sweep.

NO MATCH/NOT EXEMPTION ELIGIBLE List. Records for all currently registered voters in TEAM which do not list the voter as having already applied for and received the disability exemption and for which no matching record is found in any identification or disability database on any sweep.

NO MATCH/NOT EXEMPTION ELIGIBLE/NOT AGE VOTE-BY-MAIL ELIGIBLE List. Records for all currently registered voters in TEAM which do not list the voter as having already applied for and received the disability exemption, for which no matching record is found in any identification or disability database on any sweep, and which do not establish that the voter is qualified to vote by mail on account of age.

B. Data Used in the Matching Process

State of Texas Databases

25. Counsel for the Department of Justice provided me with voter registration and identification databases from the State of Texas.

26. Texas Voter Registration Data: Texas voter registration records were extracted from the TEAM system on January 15, 2014. That date serves as the date of the election for purposes of this analysis. Any form of acceptable SB 14 photo ID that is unexpired on that date or expired no earlier than sixty days prior to that date is deemed valid under SB 14 for purposes of this analysis.

27. State of Texas Photo ID Data: Records were extracted from the Texas Department of Public Safety (DPS) databases for Driver Licenses (DL), Personal State Identification Cards (PID), Licenses to Carry Concealed Handguns (CHL), and Election Identification Certificates (EIC). The DPS DL, PID, CHL, and EIC databases also include a field that indicates whether an individual may be deceased. The algorithm matches TEAM records to identification records for each specific form of ID separately.

Federal Databases

28. I created a version of the TEAM database with identifiers used in matching already formatted to facilitate matching to Federal databases. I developed a step-by-step explanation of the algorithm and model computer code in STATA and SQL to guide the matches of the Federal databases. Counsel for the United States conveyed these files to staff at the United States Citizenship and Immigration Service, the United States Department of Defense, the United States Department of State, the United States Department of Veterans Affairs, and the Social Security Administration. I had no direct access to Federal databases.[7]

C. Databases Used in Data Analyses

29. I analyze individual and aggregate-level data to estimate the relationship between race and probability of possession of ID accepted under SB 14. Individual-level analyses rely on TEAM data matched to identification and disability databases and Catalist racial classifications.

30. Aggregate-level analyses in Section VI (Table VI.1) sum the number of Matched and Not Matched voters in each Census Block Group (BG). I calculate the percent of registered voters in each BG that are on the NO MATCH list. Block Group is the lowest level of aggregation for which Citizen Voting Age Population (CVAP) is

[7] Staff at each agency involved in the matching process completed a declaration documenting the steps that they took in completing the matches. Those declarations are provided as exhibits in the Appendix.

released by the Census. I correlated the percent NO MATCH with the percent of each racial group that are adult citizens (CVAP) or are adults (VAP) at the BG level. The correlation allows me to estimate the share of the Citizen Voting Age Population (CVAP) or Voting Age Population (VAP) for each racial group that lacks acceptable SB 14 photo ID. The correlations and ecological regressions reported in Section VI are of the NO MATCH and MATCH data aggregated to the BG level for the CVAP.[8] Similar results hold for VAP, but CVAP is a closer approximation to the eligible electorate. I also tested the robustness of the analysis at the Census Tract level (a higher level of aggregation than BG), and found no substantive or statistical difference in the results. I report results at the BG level.

31. Aggregate analyses in Section VII on rates of turnout and registration are at the Voting Tabulation District (VTD) level as that is the level at which registration and voting statistics are reported. These analyses use Voting Age Population (VAP) from the Census Enumeration and Citizen Voting Age Population (CVAP) from ACS.[9]

Voting and Registration Data

32. Analyses of historical voting and registration patterns use aggregate data from the Texas Legislative Council, a nonpartisan State legislative agency. Voting and

[8] In the case of VAP, I used the Census figures for the VAP for each group at the BG level. In the case of CVAP, I use ACS data aggregated to the BG.
[9] For Block Groups that contain multiple VTDs, I apportion the CVAP of a racial group in a Block Group based on the percent of the VAP of that group in a Block Group that resides in each VTD.

registration data at the VTD level come from the website of the Texas Legislative Council.[10] Individual data on voter participation history in particular elections is also used from the TEAM database.

Census Data

33. The Population and Voting Age Population in Voting Tabulation Districts (VTDs) are collected by the 2010 Census Enumeration and come from the website of the Texas Legislative Council.[11]

34. Aggregate data on the Citizen Voting Age Population (CVAP), at the Census Block Group and VTD level, come from the five-year average of the American Community Survey (ACS), 2008-2012.[12]

35. The Census Bureau conducts the Voting and Registration supplemental to the CPS each November following federal elections. I examine the 2006, 2008, 2010, and 2012 CPS.[13]

[10] ftp://ftpgis1.tlc.state.tx.us/elections/
[11] ftp://ftpgis1.tlc.state.tx.us/2011_Redistricting_Data/VTDs/
[12] *See* http://www.census.gov/acs/www/data_documentation/2012_release/
[13] http://www.census.gov/hhes/www/socdemo/voting/publications/p20/index.html

Catalist Database

36. The United States contracted with Catalist, LLC, to obtain additional data on voter registration records in the State of Texas. I use Catalist data (i) to test the robustness and validity of findings, (ii) to obtain estimated classifications of the race of individuals (estimates that are based in part on local area Census demographics and frequencies of names), (iii) to examine voting history of individuals, and (iv) to obtain geocoding of each registration record and Catalist deadwood, deceased, and NCOA information.

37. Catalist maintains data on voter registrations and vote history. Catalist retains the TEAM VUID, which permits linkage between the Catalist data and the TEAM data.

38. Catalist augments official voter registration data with indicators regarding whether a voter has moved as reported through the National Change of Address (NCOA) data from the United States Postal Service, information on whether a voter is deceased using data from the Social Security Administration and private vendors, and information identifying potentially obsolete records (or deadwood) based on factors including participation in past elections. To analyze the validity and robustness of findings, I use the Catalist flags for NCOA, deceased, and deadwood records to construct subsets of the pool of registration records.

39. Catalist provides a classification of the race of each individual in the TEAM database and a score for the confidence in that classification. Catalist's race classification is based on the frequency of specific last names in the population and the frequency of racial groups in local areas (Census block groups), and is a refinement on the area-based estimates underpinning the ecological regression analysis frequently used in Voting Rights Act cases.

40. Catalist data are widely used in academic research on registration and voting and have been vetted for publication in peer-reviewed journals. Catalist data on demographic characteristics of the electorate, including age, gender, and race, have been vetted and published in peer-reviewed journals.[14] Academic researchers use the Catalist database to identify the population of registered voters in the US and to conduct random sample surveys of the population of registered voters and experimental research on voter participation.[15]

---

[14] See for example, Ansolabehere, Stephen, and Eitan Hersh. "Validation: What big data reveal about survey misreporting and the real electorate." *Political Analysis* (2012): mps023. Ansolabehere, Stephen, Eitan Hersh, Kenneth Shepsle. "Movers, Stayers, and Registration: Why Age is Correlated with Registration in the US." *Quarterly Journal of Political Science* 7, no. 4 (2012): 333-363.Ansolabehere, Stephen, and Eitan Hersh. "Gender, Race, Age and Voting: A Research Note." *Politics and Governance* 1, no. 2 (2013): 132-137. Garcia-Castañon, Marcela, Alison D. Rank, and Matt A. Barreto. "Plugged in or tuned out? Youth, race, and Internet usage in the 2008 election." *Journal of Political Marketing* 10, no. 1-2 (2011): 115-138.

[15] See, for example, Nickerson, David W., and Todd Rogers. "Political Campaigns and Big Data." *The Journal of Economic Perspectives* 28, no. 2 (2014): 51-73. Dale, Allison, and Aaron Strauss. "Don't forget to vote: Text message reminders as a mobilization tool." *American Journal of Political Science* 53, no. 4 (2009): 787-804. Bennion, Elizabeth A., and David W. Nickerson. "The Cost of Convenience An Experiment Showing E-Mail Outreach Decreases Voter Registration." *Political Research Quarterly* 64, no. 4 (2011): 858-869. Ansolabehere, Stephen, and Eitan Hersh. "Validation:

41. I verify the validity of the inferences for the full Catalist analysis by examining the subset of records in Catalist with very high confidence in the racial classification. Catalist has very high confidence in areas where there is a fairly homogeneous population and for individuals with sufficiently distinctive names.

42. In addition to the analysis using Catalist data, I analyze Census data on the racial composition of the electorate and the rate with which registered voters are deemed to possess acceptable SB 14 identification. I perform Ecological Regression at the Block Group (BG) level showing how the rate of NO MATCH corresponds with the racial composition of the electorate. The Ecological Regression is completely consistent with the analysis of the racial disparities on the NO MATCH list using Catalist's racial classification estimates. The Ecological Regression, then, provides independent evidence of racial disparities in possession of SB 14 IDs and confirmation of the validity of inferences drawn using the Catalist data.

V. Matching Process

A. Description of the Matching Process

43. The matching process conducts record linkage for individual records in TEAM to any record in each database corresponding to a form of identification accepted

---

What big data reveal about survey misreporting and the real electorate." *Political Analysis* (2012).

under SB 14 or in a database reflecting those eligible to apply for the disability exemption. The matching algorithm proceeds in four parts.

1. Database Preparation. Databases are prepared and standardized.
2. Creation of Identifiers. Identifier values used to link records in one database to records in another database are constructed by combining multiple individual fields.
3. Record Linkage and Matching. One-to-many matches are conducted between the databases. That is, the algorithm matches each unique identifier on the TEAM database to all records on the identification database that have the corresponding value of the identifier.
4. Data Gathering. Appended to the TEAM data are fields indicating every match found of a record on the TEAM database to a record on a state or Federal identification database or Federal disability databases.

44. Each of the four parts is divided into multiple stages, which are in turn divided into concrete steps. Detailed procedures for the implementation of the algorithm were prepared and presented to all parties in the litigation and to the Federal agencies. These memorandums provide detailed, step-by-step documentation of the matching process. Rather than reproduce the detailed steps here, the memorandums presented to all parties and the Federal agencies are appended to this report in the appendix. This section describes the stages of the Matching Process at a general level.

45. The result of this methodology is to produce a MATCH list, a NO MATCH list, and a NO MATCH/NOT EXEMPTION ELIGIBLE list, as described above.

B. Features of the Matching Algorithm

46. The first phase of the matching algorithm, Database Preparation, standardizes the coding of database fields to facilitate matching. Different databases store the fields in different ways. For example, Gender is coded 1 or 0 in some databases and M or F in others. The database preparation in the algorithm standardizes the coding of names by removing spaces, hyphens, and other characters; standardizes dates of birth and gender codes, and identifies invalid or missing values (such as 111111111 for Social Security Numbers). Prior research has shown that standardization of fields, removal of duplicate cases, and definition of missing or invalid values in each field greatly improves the quality of matching.[16]

47. The second part of the algorithm develops multiple identifiers for purposes of record linkage. This general approach is widely used in the field of record linkage and database matching, especially in health research and marketing, and has been

[16] William E. Winkler, "Methods for Evaluating and Creating Data Quality," Information Systems 29 (October, 2004), 531-550.
http://www.sciencedirect.com/science/article/pii/S030643790400002X
Max G. Arellano and Gerald I. Weber, "Issues in Identification and Linkage of Patient Records Across an Integrated Delivery System," Journal of Health Care Information 12, no. 3 (1998): 43-52
http://sce.umkc.edu/~leeyu/Mahi/medical-data5.pdf

determined by past studies to yield a very high rate of correct matches.[17] The algorithm builds identifiers by combining fields related to Address, Date of Birth, Gender, Name, Social Security Number, and Texas Driver License Number.

49. In total 13 different identifiers were constructed in the TEAM database and in the corresponding State and Federal databases.[18] Each identifier corresponds to a particular combination of fields. For example, Combination A consists of First Name, Last Name, Date of Birth, Gender, Street Number, and 5-digit ZIP Code. A sample version of Combination A for a man named John Smith, born on January 1, 1960, and living at 100 Main Street in the ZIP Code 78610 would be JOHNSMITH01011960110078600. A chart of the identifiers created for matching is provided below in Table V.1. Appendix Tables A.V.1-A.V.2 provide statistics on the completeness and uniqueness of combinations of fields.

[17] See Simon, Michael S., Beth A. Mueller, Dennis Deapen, and Glenn Copeland. "A comparison of record linkage yield for health research using different variable sets." *Breast cancer research and treatment* 89, no. 2 (2005): 107-110. Sweeney, Latanya, "Computational Disclosure Control for Medical Microdata: The Datafly System, Record Linkage Techniques 1997, Chaper 11. Pp. 442-453. Sweeney, Latanya, Matching Known Patients to Health Records in Washington State Data. Harvard University. Data Privacy Lab. 1089-1. June 2013.

[18] For the Department of State matches of passport holders, additional identifiers were developed to address specific features of the way that the Passport database stores name information. Those additional identifiers are discussed in a declaration from DOS, which is attached in the appendix.

Table V.1. Combinations of Fields Used as Matching Identifiers	
Combination Code	PRIMARY MATCHES
A	First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number
B	Last Name + Gender + DOB + Residential ZIP + Residential Street Number
C	Gender + DOB + Residential ZIP + Residential Street Number
D	First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number
E	First Name + Last Name + Gender + Residential ZIP + Residential Street Number
F	First Name + Last Name + Gender + DOB
M	Texas Driver License Number (where available)
	SECONDARY MATCHES
G	First Name + Middle Initial + Last Name + DOB
H	Last 4-Digit SSN + DOB + Residential ZIP
I	Last 4-Digit SSN + First Name + Last Name + DOB
K	First Name + Last Name 1 + Middle Initial + DOB
L	First Name + Last Name 2 + Middle Initial + DOB
SSN	9-Digit Social Security Number

50. The third stage of the process, the Record Linkage and Matching phase, conducts one-to-many matches[19] and performs multiple sweeps to guard against

[19] As used here, a "one-to-many" match means that for each TEAM record, a match will only be attempted for a particular identifier when that identifier is unique to a single TEAM record. However, that unique identifier will be deemed to have matched to any identical identifier in an identification or disability database, even if the identifier is not unique on the identification or disability database. As an example, consider a simplified identifier not used here: First Name + Last Name + Date of Birth. If there are two people named JOHN SMITH with a birth date of January 1, 1960 in TEAM, no match is attempted on First Name + Last Name + Date of Birth because the TEAM identifier is not unique. This guards against false positives. On the other hand, if there is only one JOHN SMITH in TEAM born on January 1, 1960, and there are two JOHN SMITHs born on January 1, 1960 in the DPS driver license database, the match will be attempted, and the TEAM record is considered to have matched. All that matters is that the registered voter John Smith

false negatives (non-matches that should be matched). An example of false negatives that this approach guards against is a typographical error in the spelling of a person's name as recorded on TEAM but not in the DPS driver license database. Such a typo would create an inconsistency between TEAM and DPS if matches were conducted only on identifiers that included name. In addition to searching on identifiers that contain name fields the algorithm searches on identifiers constructed from combinations that do not include name elements, such as Date of Birth, Gender, Address, and Social Security Number. The algorithm will match the record on the identifiers that do not contain each of these categories of fields, thus avoiding non-matches due to typographical errors, nicknames, missing fields, and other inconsistencies between databases. A record is determined to have found a match if a given identifier in TEAM is identical to at least one corresponding identifier in an identification database. The frequencies of matches of individual records in TEAM to specific identifiers in each of the state and federal identification and disability databases are presented in the appendix. See Appendix Tables A.V.3 to A.V.5.

51. As shown in the chart below, the algorithm conducts two sorts of sweeps through the data to find matching records. The Primary Sweeps match on Combinations A-F and M, and are run on all TEAM records. The Secondary Sweeps are conducted on the TEAM records not matched in the Primary Sweeps for Combinations G – L. For Federal databases, the Primary Sweeps are run against all

---

is deemed to hold a driver license—there is no need to distinguish between the two DPS records.

qualifying Federal records with Texas addresses, while the Secondary Sweeps are run both against Texas-only records, as well as against the nationwide universe of the relevant Federal dataset.

		Matching Combinations
Texas DPS Databases	**Primary Sweeps** (All TEAM records)	Combination A: First name + Last name + Gender + DOB + Street number + ZIP Combination B: Last name + Gender + DOB + Street number + ZIP Combination C: Gender + DOB+ Street number + ZIP Combination D: First name + Last name + Street number + ZIP Combination E: First name + Last name + Gender + Street number + ZIP Combination F: First name + Last name + DOB + Gender Combination M: Texas Driver License Number
	Secondary Sweeps (TEAM records with no primary match)	Combination G: First name + Last name + Middle Initial + DOB Combination H: DOB + ZIP + SSN4 Combination I: First name + Last name + DOB + SSN4 Combination K: First name + Last name 1 + Middle Initial + DOB Combination L: First name + Last name 2 + Middle Initial + DOB[20] SSN: 9-digit Social Security Number
Federal Identification and Disability Databases	**Primary Sweeps** (All TEAM records against Federal records with a Texas address)	Same as primary sweeps for DPS databases, except for Texas Driver License Number (Combinations A-F)
	Secondary Sweeps (TEAM records with no primary match against Federal records with a Texas address)	Same as secondary sweeps for DPS databases (Combinations G-L and SSN)
	Nationwide Sweeps (TEAM records with no primary or secondary match against nationwide Federal records)	All sweeps without address criteria (Combinations F, G, I, K, L, and SSN)

[20] "Last name 1" is the first half of a hyphenated last name, and "Last name 2" is the second half of a hyphenated last name. Combinations K and L in TEAM are each matched against Combination G, Combination K, and Combination L in the identification and disability databases for a total of six matching sweeps.

52. In the last phase of the matching process, the Data Gathering phase, the results of all matching sweeps are recorded for each individual TEAM record. This stage also appends indicators of deceased records from the Texas DPS data to TEAM.

53. The algorithm developed for DOJ in the present case goes beyond the matching algorithm developed for the expedited Section 5 proceedings in *State of Texas v. Holder* in which I also testified on behalf of the United States.[21] First, the TEAM database is now matched to all relevant state and federal databases. In the section 5 proceeding, time constraints prevented an assessment of federal identification and disability data. Second, the algorithm ensures that persons that DPS considers to be dead are not included on the NO MATCH list. Instead, deceased voters are identified after matching has occurred.[22] Third, by using multiple identifiers, the algorithm is developed to be sensitive to variations in names, such as nicknames and compound names, to typographical errors, and to missing information. Fourth, by matching on identifiers constructed from a larger number of categories of fields (three or four rather than two), the algorithm is more precise in determining which records link and more exhaustive in the search for linkages.

54. The data analysis presented in the Validation section of this report examines the robustness of results to alternative racial classifications.

---

[21] The Court in *Texas v. Holder* chose not to rely on any of the expert testimony presented. 888 F. Supp 2nd 113, 134-138 (D.D.C. 2012).

[22] Where a TEAM record has matched to a DPS record that the State of Texas has marked as deceased, recording this information allows matched records for dead people to be excluded from further analysis of the population of voters actually affected by SB 14's requirements.

C. MATCH , NO MATCH and NO MATCH/NOT EXEMPTION ELIGIBLE Lists

Match Rates

55. The implementation of the algorithm developed for the United States in this case matched the entire TEAM database to 10 different state and federal databases. Table V.2 below lists the number of records in TEAM that matched to each state or Federal database using that algorithm, as well as the percent of TEAM records overall and in each racial group that match to each identification or disability database.

56. The most commonly held form of identification is a State of Texas Driver License, followed by a United States Passport. Just over 78 percent of records in TEAM matched to the DPS Driver License list, while 42 percent of records in TEAM matched to the DOS passport database. The next most common form of ID is a DPS Personal (or State) ID, held by 10 percent of those in TEAM.

57. Table V.2 shows that the rates with which records match to the databases varies with race. The percentage of records on TEAM matched to Driver Licenses and Passports is much higher for Anglos than for Blacks and Hispanics. 83 percent of Anglos on TEAM match to a record on the DPS Driver License database, compared with 67 percent for Blacks and 72 percent for Hispanics. 46 percent Anglos on

TEAM match to a record on the DOS passport database, compared with 25 percent for Blacks and 38 percent for Hispanics.

Table V.2. Number of Matches of TEAM Records to State and Federal Databases Overall and By Racial Group, using Catalist Racial Estimates
(Percent of TEAM Records that Match to a Given ID or Disability Database)

Database	Race				
State of Texas ID Databases	Anglo	Black	Hispanic	Other	All
Driver License	6,894,742 (83.1%)	1,154,170 (67.1%)	2,199,956 (72.0%)	414,870 (84.2%)	10,663,738 (78.6%)
Personal ID	443,120 (5.3%)	309,063 (18.0%)	502,031 (16.4%)	30,444 (6.2%)	1,284,658 (9.5%)
Concealed Handgun License	588,087 (7.1%)	57,129 (3.3%)	72,953 (2.4%)	14,839 (3.0%)	733,008 (5.4%)
EIC	69	43	51	0	163
Federal ID Databases					
DOS	3,776,207 (45.5%)	424,682 (24.7%)	1,151,608 (37.7%)	378,666 (76.8%)	5,731,163 (42.3%)
DOD	427,191 (5.2%)	81,688 (4.8%)	116,460 (3.8%)	13,015 (2.6%)	638,354 (4.7%)
USCIS	106,051 (1.3%)	45,005 (2.6%)	373,576 (12.2%)	210,454 (42.7%)	735,086 (5.4%)
VHA	186,695 (2.3%)	49,179 (2.9%)	57,635 (1.9%)	2,496 (0.5%)	296,005 (2.2%)
Federal Disability Databases					
SSA: Disability	419,065 (5.1%)	167,980 (9.8%)	202,368 (6.6%)	14,925 (3.0%)	804,338 (5.9%)
VBA: Disability	118,883 (1.4%)	31,952 (1.9%)	35,743 (1.2%)	1,938 (0.4%)	188,516 (1.4%)

58. The rate at which records on TEAM match to DPS Personal ID or USCIS databases is higher for minorities than for Anglos. Of Anglos on TEAM, 5 percent

match to records on the DPS Personal ID database and 1 percent match to the USCIS database of holders of certificates of citizenship and naturalization. Of Blacks on TEAM, 18 percent match to the Personal ID database and 3 percent match to the USCIS database. Of Hispanics on TEAM, 16 percent match to the Personal ID database and 12 percent match to the USCIS databases. However, both the DPS Personal ID and USCIS Certificates of Citizenship and Naturalization are less commonly held than Driver Licenses and Passports.

Sizes of the MATCH, NO MATCH and NO MATCH/NOT EXEMPTION ELIGIBLE lists

59. Of the 13,564,416 records in the TEAM database, 11,527,531 matched to at least one record corresponding to acceptable SB 14 photo ID issued by the State of Texas, and 6,326,128 records matched to at least one record corresponding to acceptable SB 14 photo ID issued by the Federal government. Most of the records matched to the Federal databases also matched to a State of Texas identification database. No matching record was found on any of the state or Federal identification databases and no disability exemption was grated for 1,232,242 records on the TEAM database, approximately 9.1 percent of all records in TEAM. No matching record was found for 1,117,496 records in the TEAM database matched to a state or Federal identification database or to a Federal disability database. Tables V.5 and V.6 summarize the results of the matching process.

60. Removing matched voters that DPS data indicate as deceased from the number of records in the TEAM database reduces the universe of registered voters from 13,564,416 to 13,515,671. I define this set, which removes from further analysis voters that the State of Texas data from DPS indicate to be deceased, as the Baseline Universe of Registered voters.

VI. Results: Analysis of Racial Disparities

61. This section analyzes how the rate of non-matched records between TEAM and valid SB 14 ID databases varies across racial groups.

A. Census Racial Data and Possession of ID

62. The American Community Survey (ACS) conducted by the Census Bureau provides estimates of the racial composition of the electorate. Using the 5-year average of the survey from 2008 to 2012, the ACS provides estimates of the CVAP of various racial groups at the block group level and tract level.

63. I aggregated the Match and No Match lists to the block group level, the lowest geographic level at which Census reports ACS CVAP numbers.[23] Within each block group, I computed counts of the numbers of registered voters and the number of

[23] I conducted similar analyses at the Census tract level and discovered the same pattern of results. Because Block Group is a lower level of aggregation I present that here.

registered voters who did not match any identification database and the number of registered voters who were matched to at least one identification database. I then computed the percentage of registered voters in each block group who were not matched to any record.

64. I used ecological regression and homogeneous block group analyses[24] to estimate the percentage of Black citizens of voting age, Hispanic citizens of voting age, and Anglo citizens of voting age for whom a matching record to an identification database was found. Ecological regression estimates the relationship between Percent No Match and Percent of CVAP who are Anglo, Black, or Hispanic, enabling me to estimate the percent of each group who match to an ID database. Homogeneous block group analysis examines the subset of block-groups where all or almost all of the adult citizens are of one race—in this analysis, at least 80 percentage of a given race. Within such racially homogeneous areas, I then compute the percent of registered persons for whom a matching identification record was found or not found; this enables me to estimate the percentage of that group who has (or does not have) an SB 14 ID. Both ecological regression analysis and homogeneous block group analysis are well-established statistical procedures relied upon in voting rights cases, where they are often used to measure racially polarized voting and cohesiveness of voting of racial groups.

[24] In other contexts data are at the precinct level, so homogeneous block group analysis is also called homogeneous precinct analysis.

65. Table VI. 1 presents the ecological regression and homogenous block group analysis estimates of the rate with which No Match was found among each racial group. The ecological regression estimates indicate that no match between TEAM and a state or Federal identification database was found for approximately 5 percent of Anglos, 13 percent of Blacks, and 9 percent of Hispanics. The difference between the Black and Anglo rate of non-matching, then, is 8 percentage points, and the difference between Hispanics and Anglos is approximately 4 percentage points. In other words, Hispanics are almost twice as likely as Anglos to be on the NO MATCH list, and Blacks are approximately two and a half times as likely as Anglos to be a NO MATCH.

66. The homogeneous block group analyses in Table VI.1 are similar to the ecological regression estimates. Almost 7 percent of Anglo registered voters were estimated to have no matching identification record on a state or Federal database. That figure was nearly 20 percentage points among Blacks and 14 points among Hispanics. In other words, the difference in the rate of NO MATCH equaled 13 points between Blacks and Anglos and 7 points between Hispanics and Anglos. In the homogeneous blocks, Hispanics are twice as likely as Anglos and Blacks are three times as likely as Anglos to not match to an appropriate identification database.

67. The results from the analysis of the homogeneous block groups and the ecological regressions are highly unlikely to have arisen by chance. The observed

differences across the groups are statistically significant at the confidence levels generally used by social scientists.

B. Catalist Racial Data and Possession of ID

68. Analysis of individual level data using the Catalist classification of race yields similar results to the aggregate analyses presented in part VI.A. The Baseline Universe of Registered Voters, which consists of all currently-registered voters in TEAM – save for those who matched a record marked as deceased in a DPS ID file – has 13,515,671 records. Of these, 8,263,528 are classified as Anglo according to Catalist's estimates; 1,712,388 are Black; 3,047,970 are Hispanic; and 491,785 are Other Races.

69. The rate of non-matches between TEAM and identification databases varies by race. Of records identified as Anglo in the Baseline Universe of Registered Voters, 7.4 percent had no matching record in state or Federal identification databases. By comparison, no matching records were found for 15.1 percent of people identified as Black and 11.3 percent of people identified as Hispanic. See Table VI.2.

70. The differences in rates of matching and non-matching across racial groups are statistically significant at the confidence levels normally used by social scientists. The difference between Blacks and Anglos in the rate of non-matching is 7.7 percentage points. Such a difference is highly unlikely to have arisen by chance

given the degree of precision of the estimates.[25] Likewise, the difference between Anglos and Hispanics in the rate of non-matching is 3.9 percentage points, which is highly unlikely to have arisen by chance.[26]

C. Eligibility for Exemption and Vote-by-Mail

71. The analysis in Table VI.2 does not account for the fact that those with a Federally-determined disability may qualify and apply for an exemption. Also, those over 65 may vote by mail without showing ID, but must still show ID if they vote at a polling place.

72. Table VI.3 parallels the analysis in Table VI.2, but classifies as MATCHES records for which an exemption is possible. The first column of the table shows the relationship between Race and percent on the NO MATCH/NOT EXEMPTION ELIGIBLE. That list consists of those not matched to a state or Federal identification database or to a Federal disability database. The second column of the table shows the relationship between Race and percent on the NO MATCH/NOT AGE VOTE-BY-MAIL ELIGIBLE. That list consists of those not matched to a state or Federal identification database and under 65 years of age. The third column of the table shows the relationship between Race and percent on the NO MATCH/ NOT EXEMPTION ELIGIBLE/NOT AGE VOTE-BY-MAIL ELIGIBLE. That list covers people

[25] The 99 percent confidence interval for that difference in proportions is +/- 0.1 percentage points, so the differences are statistically different from 0.
[26] I use a 99 percent confidence level. That is the probability of observing a difference this large by chance is less than 1 percent.

not potentially exempt for reasons of disability or for reasons of age and not matched to a state or Federal ID. For this analysis the Baseline of 13,515,671 records is used as the pool of registered individuals.

73. Considering potential disability and age exemptions, 6.8 percent of the registration records in TEAM were NO MATCH/NOT EXEMPTION ELIGIBLE/NOT AGE VOTE-BY-MAIL ELIGIBLE. Among records classified as Anglo by Catalist's estimates, 5.3 percent were not matched to any record in any identification or exemption database and were under 65. By comparison, a NO MATCH and no potential exemption were found for 11.5 percent of people classified as Black and 9.0 percent of people classified as Hispanic. Similar patterns hold for each potential exemption (disability and age) separately. See Table VI.3. These estimates are very similar to those in Table VI.2.

74. The difference between Black and Anglo rates for having neither a MATCH nor an exemption is 6.2 percentage points, and the difference between Hispanic and Anglo rates of NO MATCH and no exemptions is 3.7 percentage points. Both differences are statistically significant at the confidence levels generally used by social scientists.

75. The aggregate and individual-level data are remarkably consistent. They show statistically significant differences between the rate with which Anglos and the rates with which Blacks and Hispanics on the TEAM database are matched to state and

federal identification databases or are eligible for an exemption. The difference between Blacks and Anglos is in the range of 7 to 8 percentage points and the difference between Hispanics and Anglos is in the range of 3 to 4 percentage points.

D. Voting Rate among Registrations with No Match to an ID

76. SB 14 affects individuals who voted in the most recent federal elections. Of all registered voters on the NO MATCH list 34.6 percent are recorded by TEAM as having voted in the 2012 election and 17.6 percent voted in the 2010 election. Of all records on the NO MATCH/NOT EXEMPTION ELIGIBLE list, 34.7 percent voted in 2012 and 17.7 percent voted in 2010.

77. The NO MATCH list contains 5.4 percent of all registered voters on TEAM who voted in 2012 and 4.7 percent of all registered voters who voted in 2010. In other words, SB 14 may impose requirements for identification that are not currently met by approximately 5 percent of all registered persons who voted in recent federal general elections.

78. Table VI.4 shows that this effect on the electorate varies by race. In 2012 approximately 4.3 percent of Anglos who were registered and voted could not be matched to identification databases. By comparison, 10.8 percent of Blacks and 6.4 percent of Hispanics who were registered and voted could not be matched to identification databases. In other words, Blacks who voted in recent federal

elections are twice as likely to lack appropriate ID under SB 14 as Anglo voters, and Hispanic voters are 50 percent more likely to lack appropriate ID.

VII. Validation

79. This section determines the sensitivity of results to alternative specifications of the pool of registered voters and to alternative classifications of race. I also examine the overlap of the NO MATCH lists produced by the algorithms of the United States and of the State of Texas.

A. Variations in Universe of Registered Voters

80. The TEAM database is the official list of registered individuals in the State of Texas. Even so, there may be questions raised about the currency of some records in that database. All states that have voter registration will have at least some registrations that are out-of-date or invalid but still on the rolls, for a wide variety of reasons. I constructed alternative formulations of the list of registered voters to examine whether such concerns could affect the inferences drawn regarding racial disparities in rates of acceptable SB 14 ID possession.

81. This section examines variations in the pool of registered voters achieved by omitting records from TEAM for which there is some indication that the record may no longer be current or valid according to information from Catalist or internal to

the TEAM database. Omitting these cases changes both the numerator and the denominator of calculations of the percent who are matched or not.

82. Table VII.1 presents the rates with which racial groups match to the relevant state and federal databases under different constructions of the set of registered voters.

Catalist Deceased, Deadwood, and NCOA flags

83. The Catalist database includes indicators of whether an individual is deceased, is deadwood (an obsolete record), or has a National Change of Address application on file with the U.S. Postal Service, indicating that the individual has moved.[27] Each of these categories provides evidence that a given registration record may no longer be current. Column 1 in Table VII.1 presents the overall NO MATCH figure and the rate of NO MATCH among racial groups excluding records flagged by Catalist as deceased, deadwood, or NCOA are excluded from TEAM.

84. Overall, 9.0 percent of the registration records in this set of registered voters were not matched to a record in a state or federal identification database. Among records that the Catalist race estimates classified as Anglo, 7.3 percent were not matched to any record in any identification database. By comparison, no matches

[27] NCOA flags do not distinguish between in-county, in-state, and out-of-state moves.

were found for 15.0 percent of people classified by Catalist as Black and 11.1 percent of people classified as Hispanic.

85. The differences between the racial groups are statistically significant. The difference between Black and Anglo rates of no matches is 7.7 percentage points, and the difference between Hispanic and Anglo rates of no matches is 3.8 percentage points. Both differences are highly unlikely to arise by chance.

Suspense Voters

86. The TEAM database distinguishes Active and Suspense (or inactive) voters. A suspense voter is still legally registered but may be dropped from the registration list for reasons of non-voting or non-response to election office communications.[28] Column 2 in Table VII.1 presents the overall NO MATCH figure and the rate of NO MATCH among racial groups excluding Suspense Voters from the pool of Registered Voters.

87. Overall, 8.3 percent of the registration records in this set of registered voters were not matched to a record in a state or Federal identification database. Among records that Catalist estimates classified as Anglo, 6.6 percent were not matched to

[28] The Suspense List, as described in Section 15.081 of the Texas Election Code, is maintained by the voter registrar in each county. It contains the names of (1) voters that failed to respond to a confirmation notice, (2) voters whose renewal certificate was returned to the registrar as undeliverable, and (3) those individuals that were excused or disqualified from jury service because they were not a resident of a given county.

any record in any identification database. By comparison, no matches were found for 14.2 percent of people classified as Black and 10.6 percent of people classified as Hispanic.

88. The differences between the racial groups are statistically significant. The difference between Black and Anglo rates of no matches is 7.6 percentage points, and the difference between Hispanic and Anglo rates of no matches is 4.0 percentage points. Both differences are highly unlikely to arise by chance.

Expired IDs

89. The DPS database lists records with expired IDs (DL, PID, CHL, and EIC). An expiration may simply mean that the individual allowed the driver license to expire because that person no longer drives. However, an expired license may also signal that the individual is no longer at a given residence.[29] Column 3 in Table VII.1 presents the overall NO MATCH figure and the rate of NO MATCH among racial groups excluding from the pool of Registered Voters those who could be matched to DPS records with Expired IDs. It should be noted that in order to extract the expiration information, I first matched the records in TEAM to the DPS files and then omitted from the analysis all records that had expired IDs.

[29] Therefore, exclusion of these records removes some registered voters who no longer reside at the residence at which they are registered, and perhaps not in the State of Texas. It also removes records of some people who remain at their residence but allowed their ID to expire and will be affected by SB 14.

90. Overall, 6.7 percent of the registration records in this set of registered voters were not matched to a record in a state or federal identification database. Among records that the Catalist estimates classified as Anglo, 5.3 percent were not matched to any record in any identification database. By comparison, no matches were found for 11.6 percent of people identified as Black and 8.5 percent of people identified as Hispanic.

91. The differences between the racial groups are statistically significant. The difference between Black and Anglo rates of no matches when excluding expired IDs is 6.3 percentage points, and the difference between Hispanic and Anglo rates of no matches is 3.2 percentage points. Both differences are highly unlikely to arise by chance.

All Filters

92. Finally, I exclude from the pool of registered voters those records with Catalist deceased, deadwood, or NCOA flags, those with Expired IDs, and those listed as Suspense Voters from TEAM. Column 4 in Table VII.1 presents the overall NO MATCH figure and the rate of NO MATCH among racial groups registrations when excluding Expired IDs, Suspense registrations, and registrations for which there is a Catalist flag for deceased, deadwood, or NCOA.

93. Overall, 6.2 percent of the registration records in this set of registered voters were not matched to a record in a state or federal identification database. Among records that Catalist estimates classified as Anglo, 4.8 percent were not matched to any record in any identification database. By comparison, no matches were found for 11.0 percent of people classified as Black and 8.0 percent of people classified as Hispanic.

94. The differences between the racial groups are statistically significant. The difference between Black and Anglo rates of no matches is 6.2 percentage points, and the difference between Hispanic and Anglo rates of no matches is 3.2 percentage points. Both are highly unlikely to have arisen by chance.

95. These analyses suggest that the general pattern of racial differences holds even under varying constructions of the pool of registered voters. It is possible to perform further analyses using combinations of different filters and screens examined here. Appendix Tables A.VI.1 and A.VI.2 present results for the disparities on the incidence of NO MATCH for the subset of TEAM records for which Catalist shows that a record may be questionable as deceased, deadwood, or NCOA.

B. Racial Classification

96. The racial classification provided by Catalist is an estimated value for each individual based on local area demographics, frequencies of names, and other

characteristics of the individuals. Errors in classification can occur. Statistical theory predicts that such errors will tend to reduce observed differences across the groups.[30] Hence, the true differences may be larger than the estimates above. It is possible to check the robustness and validity of the racial differences observed using a subset of records for which Catalist estimates of race are of the highest confidence and by using data on Spanish Surnames on the Voter files.

97. First, the pool of registered voters can be restricted to the subset of records for which Catalist has high confidence in the racial classification. Catalist provides information on the level of confidence in that predicted value. There are 6,786,387 such records on TEAM.

98. Table VII.2 presents the numbers and percentages of each racial group in the NO MATCH and MATCH list starting with the Baseline Universe but retaining from TEAM only those records for which Catalist reports high confidence of the individual's race. Overall, 8.1 percent of records in this subset of the TEAM baseline universe did not match to a corresponding record on a state or federal identification database.

99. Racial differences in rates of NO MATCH are somewhat more pronounced in the subset of registered voters for which Catalist reported that the racial classification estimate was "highly likely" correct. Of people that Catalist's estimates classified as

[30] William G. Cochran, "Errors of Measurement in Statistics," *Technometrics* 10 (Nov. 1968): 637-666.

Anglo, 6.7 percent had NO MATCH. By comparison, 17.5 percent of people classified as Black and 11.0 percent of people classified as Hispanic on TEAM could not be matched to a corresponding record on any identification database. The percent NO MATCH is 4.3 percentage points higher for Hispanics than for Anglos, and the percent NO MATCH is 10.8 percentage points higher for Blacks than for Anglos. These differences are highly unlikely to have arisen by chance and are somewhat larger than those exhibited in Tables VI.1.

100. Second, the TEAM database includes an indicator of Spanish Surname Voter Registration (SSVR). I examine differences in NO MATCH rates for those with SSVR and those without to gain further evidence of a racial difference. Comparison of SSVR with non-SSVRs will understate the differences between Hispanics and Anglos, because the pool of non-SSVRs consists of both Anglos and Blacks.

101. Table VII.3 presents the incidence of SSVR and non-SSVR in the NO MATCH and MATCH list for the Baseline. Of Spanish Surname Voter Registrations in TEAM, 11.5 percent had NO MATCH. This estimate is very close to the estimate of 11.3 percent NO MATCH in Table VI.2 using the Catalist estimates of race. Of those people classified as Non-SSVR, 8.4 percent failed to match to an identification database. The difference between the rate of non-matches for SSVR and non-SSVR is 3.1 percentage points, and is highly unlikely to have arisen by chance. The differences between SSVR and non-SSVR and the similarity of the NO MATCH rate of

the SSVR and using the Catalist estimates for Hispanics offer further validation of the results using the Catalist data.

C. Comparison with Matching Methodology Proposed by State of Texas

102. The State of Texas's algorithm matches records using combinations of Name, Date of Birth, Social Security Number, and Texas Driver License Number. The Texas algorithm proceeds as follows: (1) preparation of TEAM and identification databases, (2) construct identifiers, (3) perform matches, and (4) gather data.

103. Counsel requested that I develop sample code to provide to the relevant Federal agencies to implement the Texas algorithm for the Federal matches. I followed the procedures laid out in Defendants' memorandum to develop code for the preparation of identifiers and the implementation of matches.

104. The State's matching algorithm differed from the algorithm developed for the United States in several respects. First, the State's algorithm matched any record in TEAM with a given identifier to all records in identification databases with the same identifier. It is a many-to-many match. That is, if there are three John Smith's with the same birth date in TEAM and two John Smiths with that same birthdate in the DPS DL file, all three of the John Smith's in TEAM are considered a match to both of the John Smiths in DPS. Second, the State's algorithm matches only using identifiers

that include Name and Date of Birth fields, SSN, or Texas DL Number. Address and Gender information was ignored.

105. Even though the United States' (DOJ) and Texas' algorithms differ, they yield very similar NO MATCH and MATCH lists. Only about 30,000 records from TEAM were matched by Texas but not DOJ; approximately 100,000 records were matched by DOJ but not Texas. See Table VII.5. Given the very high overlap of the NO MATCH lists generated by the United States' and Texas' algorithms any inferences about racial disparities will be very similar using either list.

VIII. Historical Voting and Registration Patterns

106. Registration and voting patterns in prior elections can be informative about historical patterns of behavior and election administration. To assess such historical conditions I examined three sorts of data.

107. First, I examine data from TEAM on vote history and data from Catalist on race of individuals. These data measure whether Anglo registrants voted at higher rates than Black and Hispanic registrants in 2010 and 2012.

108. Second, I examined the CPS Registration and Voting Supplement and Census reports based on the CPS of the numbers and percentages of persons and citizens of

voting age who are registered and who voted in the State of Texas and in each racial or ethnic category. I studied the reports for 2006, 2008, 2010, and 2012.

109. Third, I conducted ecological regression analyses to estimate the percentage of persons in each racial or ethnic group who are registered in the State of Texas and the percentage of registered persons in each racial or ethnic group who voted. This methodology is an established methodology for studying voting patterns in voting rights cases; see *Thornburg v. Gingles* 478 US 30, 52-54 (1986).

A. Catalist and TEAM

110. Tables VIII.1a and VIII.1b present the percentages of Registered Anglos, Blacks, and Hispanics who voted in the State of Texas in 2010 and 2012, according to the Catalist estimates of individual voter's race and TEAM data on vote history. The table presents all voters (active and suspense) in the top panel and only active voters in the bottom panel. These data show that registered Anglos voted at significantly higher rates than registered Hispanics and Blacks in the two elections immediately prior to the implementation of SB 14.

B. Ecological Regression Estimates

111. I performed ecological regressions across VTDs to estimate the registration and voting rates of various groups. For these estimates I used Census Enumeration

counts of the VAP, ACS estimates of CVAP, and Registration and Vote data reported by the Texas Legislative Council.

111. I performed two sets of ecological regressions for each of elections considered, the 2006, 2008, 2010, and 2012 elections. The first set estimates the rate of Registration as a percent of the VAP and of the CVAP for each of the groups. I regress the percent of the VAP or CVAP that is registered on the percentages of each of the groups in the VAP or CVAP. The second set of analyses estimates the voting rate of registered persons for each of the groups. This is done by regressing the percent of registered persons who voted on the percentages of each of the groups in the VAP or CVAP.

112. Tables VIII.2a and VIII.2b present the Ecological Regression estimates of the percentages of adult citizen Anglos, Hispanics, and Blacks in the State of Texas who were registered in the 2006, 2008, 2010, and 2012 elections. The 95 percent confidence interval for these estimates is reported in parentheses beneath each estimated percentage.

113. According to the ecological regression estimates in Tables VIII.2a andVIII.2b, 83 to 87 percent of Anglos of Voting Age and 84 to 88 percent of Anglo Citizens of Voting Age in Texas are registered to vote. 65 to 77 percent of Blacks of Voting Age and 75 to 80 percent of the Black CVAP are registered to vote. And, 50 to 55 percent of Hispanics of Voting Age and 75 to 80 percent of Hispanic CVAP are registered to

vote. The differences between Anglo and Hispanic registration rates and between Anglo and Black registration rates are highly unlikely to have arisen by chance (i.e., are statistically significantly different from 0).

114. Tables VIII.3a and VIII.3b present Ecological Regression estimates of the percent of registered persons who voted among specific racial or ethnic groups in Texas. According to the Ecological Regression estimates, the voting rates among registered voters of Anglo populations ranges from 10 to 20 points higher than the voting rates of Blacks, and the voting rate among registered voters of Anglo populations ranges from 15 to 30 points higher than the voting rates of Hispanic populations.

C. Current Population Survey

115. Table VIII.4 presents the CPS estimates of the percentages of adult citizen Anglos, Hispanics, and Blacks in the State of Texas who reported being registered to vote in the top panel and the percentages of each group who reported voting in the bottom panel. The margin of error of these estimates (a 95 percent confidence interval) is reported in parentheses beneath each estimated percentage.

116. The CPS estimates indicate that there existed no statistically significant difference between the percentages of Black and Anglo voter registration and turnout among adult citizens in the State of Texas from 2006 to 2012. The largest

observed difference in the CPS estimates of registration rates of Blacks and Anglos arose in 2010 and was not statistically different from 0. However, the margin of error of these estimates is very wide, 7 to 9 percentage points for Blacks, so these data do not have much statistical power.

117. The CPS estimates indicate that Hispanic adult citizens in the State of Texas are significantly less likely to be registered to vote than Anglo adult citizens in the State of Texas. In 2012, for example, the difference in these groups' registration rates is 18.5 percentage points, with a standard error of 3.7 points. The 95 percent confidence interval for this estimated difference is 7.4 points. Hence, the difference is significantly greater than 0, even though the margin of error is quite wide.

118. The CPS estimates also indicate that Hispanic adult citizens in the State of Texas are significantly less likely to vote than Anglo adult citizens in the State of Texas. In 2012, for example, the difference in Anglo and Hispanic voting rates is 22.1 percentage points, with a standard error of 7.3 points. The 95 percent confidence interval for this estimated difference is 14.6 points. Hence, the difference is significantly greater than 0.

119. Table VIII.5 presents estimates of the percent of the Registered Persons in a given group that voted. These percentages are quite similar for Blacks and Anglos, according to the CPS figures, but the rate at which Hispanic registrants vote is estimated to be 10 to 20 points lower than for Anglos.

120. The margin of error on the CPS estimates, however, is sufficiently wide that these data support a fairly wide range of possible outcomes. Specifically, one cannot reject the hypothesis that the estimates based on the CPS in Table VIII.5 are inconsistent with the results from the Catalist data in Table VIII.1. Hence, CPS does not have the statistical power to detect differences at the level found using the NO MATCH list and Catalist racial classifications. Even so, CPS shows significant differences in registration and voting between Anglos and Hispanics.

121. Overall, the Catalist figures, CPS survey estimates, and the Ecological Regression estimates show that there are statistically significant differences across racial and ethnic groups in the rate at which individuals register to vote and the rate at which they vote, given that they were already registered. The observed differences in possession of SB 14 ID, then, only add a further potential barrier to participation for groups that already have lower participation rates than Anglos in the State of Texas.

Conclusion

122. This analysis has found statistically significant and robust racial differences in the rate with which registered voters in the TEAM database fail to match to records in state and Federal databases of people with photo identification required for voting under SB 14. The results in Table VI.1 show that 5 percent of Anglos DO NOT

MATCH to applicable SB 14 identification databases, compared with 13 percent of Blacks and 9 percent of Hispanics. The results are strikingly similar in analyses relating Census racial data to the incidence of NO MATCH and in analyses relating individual level data on NO MATCH and to Catalist estimates of individual race. Those differences persist when I consider eligibility for exemption, alternative definitions of the pool of registered voters, and alternative racial classifications (such as SSVR). The observed differences imply that Black and Hispanic registered voters are significantly less likely than Anglo registered voters to possess applicable SB 14 ID or qualify for an exemption under the law.

I declare under penalty of perjury that the foregoing is true and correct. Executed this 1st day of July, 2014.

Stephen D. Ansolabehere

TABLES

Table V.1. Combinations of Fields Used as Matching Identifiers	
Combination Code	PRIMARY MATCHES
A	First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number
B	Last Name + Gender + DOB + Residential ZIP + Residential Street Number
C	Gender + DOB + Residential ZIP + Residential Street Number
D	First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number
E	First Name + Last Name + Gender + Residential ZIP + Residential Street Number
F	First Name + Last Name + Gender + DOB
M	Texas Driver License Number (where available)
	SECONDARY MATCHES
G	First Name + Middle Initial + Last Name + DOB
H	Last 4-Digit SSN + DOB + Residential ZIP
I	Last 4-Digit SSN + First Name + Last Name + DOB
K	First Name + Last Name 1 + Middle Initial + DOB
L	First Name + Last Name 2 + Middle Initial + DOB
SSN	9-Digit Social Security Number

		Matching Combinations
Texas DPS Databases	**Primary Sweeps** (All TEAM records)	Combination A: First name + Last name + Gender + DOB + Street number + ZIP Combination B: Last name + Gender + DOB + Street number + ZIP Combination C: Gender + DOB+ Street number + ZIP Combination D: First name + Last name + Street number + ZIP Combination E: First name + Last name + Gender + Street number + ZIP Combination F: First name + Last name + DOB + Gender Combination M: Texas Driver License Number
	Secondary Sweeps (TEAM records with no primary match)	Combination G: First name + Last name + Middle Initial + DOB Combination H: DOB + ZIP + SSN4 Combination I: First name + Last name + DOB + SSN4 Combination K: First name + Last name 1 + Middle Initial + DOB Combination L: First name + Last name 2 + Middle Initial + DOB[1] SSN: 9-digit Social Security Number
Federal Identification and Disability Databases	**Primary Sweeps** (All TEAM records against Federal records with a Texas address)	Same as primary sweeps for DPS databases, except for Texas Driver License Number (Combinations A-F)
	Secondary Sweeps (TEAM records with no primary match against Federal records with a Texas address)	Same as secondary sweeps for DPS databases (Combinations G-L and SSN)
	Nationwide Sweeps (TEAM records with no primary or secondary match against nationwide Federal records)	All sweeps without address criteria (Combinations F, G, I, K, L, and SSN)

[1] "Last name 1" is the first half of a hyphenated last name, and "Last name 2" is the second half of a hyphenated last name. Combinations K and L in TEAM are each matched against Combination G, Combination K, and Combination L in the identification and disability databases for a total of six matching sweeps.

Table V.2. Number of Matches of TEAM Records to State and Federal Databases Overall and By Racial Group, using Catalist Racial Estimates
(Percent of TEAM Records that Match to a Given ID or Disability Database)

Database	Race				
State of Texas ID Databases	White	Black	Hispanic	Other	All
Driver License	6,894,742 (83.1%)	1,154,170 (67.1%)	2,199,956 (72.0%)	414,870 (84.2%)	10,663,738 (78.6%)
Personal ID	443,120 (5.3%)	309,063 (18.0%)	502,031 (16.4%)	30,444 (6.2%)	1,284,658 (9.5%)
Concealed Handgun License	588,087 (7.1%)	57,129 (3.3%)	72,953 (2.4%)	14,839 (3.0%)	733,008 (5.4%)
EIC	69	43	51	0	163
Federal ID Databases					
DOS	3,776,207 (45.5%)	424,682 (24.7%)	1,151,608 (37.7%)	378,666 (76.8%)	5,731,163 (42.3%)
DOD	427,191 (5.2%)	81,688 (4.8%)	116,460 (3.8%)	13,015 (2.6%)	638,354 (4.7%)
USCIS	106,051 (1.3%)	45,005 (2.6%)	373,576 (12.2%)	210,454 (42.7%)	735,086 (5.4%)
VHA (VIC)	186,695 (2.3%)	49,179 (2.9%)	57,635 (1.9%)	2,496 (0.5%)	296,005 (2.2%)
Federal Disability Databases					
SSA: Disability	419,065 (5.1%)	167,980 (9.8%)	202,368 (6.6%)	14,925 (3.0%)	804,338 (5.9%)
VBA: Disability	118,883 (1.4%)	31,952 (1.9%)	35,743 (1.2%)	1,938 (0.4%)	188,516 (1.4%)

Table V.3. Total Records and Records for which a Match or No Match was found to Any Federal or Any State Identification Database using DOJ Algorithm*

		ANY FEDERAL RECORD		
ANY STATE RECORD		No Match to a Federal ID	Match to a Federal ID	ALL
	No Match to State ID	1,232,242	804,638	2,036,880
	Match to State ID	6,006,035	5,521,484	11,527,518
	ALL	7,238,276	6,326,122	13,564,398

* Individuals who have successfully applied for a disability exemption are counted as having matched.

Table V.4. Total Records and Records for which a Match or No Match was found to Any Federal or Any State Identification Database or any Federal Disability Database using DOJ Algorithm*

		ANY FEDERAL RECORD		
ANY STATE RECORD		No Match to a Federal ID or Disability	Match to a Federal ID or Disability	ALL
	No Match to State ID	1,117,496	919,384	2,036,880
	Match to State ID	5,533,218	5,994,300	11,527,518
	ALL	6,650,714	6,913,684	13,564,398

* Individuals who have successfully applied for a disability exemption are counted as having matched.

Table VI.1. Estimated Percent No Match By Racial Group Using Census Racial Data: Analyses of ACS CVAP and No Match Percent at Block-Group Level		
	Ecological Regression*	Homogeneous Block Groups***
Racial Group	Estimated % No Match (Margin of Error)	Estimated % No Match (Margin of Error)
Anglo	5.3% (± 0.1%)	6.9% (± 0.1%) [N of Block Groups = 4,224]
Black	13.2% (± .3%)	19.5% (± 0.4%) [N of Block Groups = 465]
Hispanic	8.8% (± .2%)	14.1% (± 0.2%) [N of Block Groups = 1554]
	Gross Percentage Point Disparity	
Black % - Anglo %	7.9%	12.6%
Hispanic % – Anglo %	4.4%	7.2%

* Number of Cases = 15,673. R^2 = .432.
** Level of analysis: Block Group;
Dependent variable: Number NO MATCH in Block Group divided by ACS CVAP Estimate in Block Group;
Multiple Regression of Percent CVAP Registered on HCVAP Percent and BCVAP Percent; Weighted by CVAP.
*** Homogeneous block groups are areas in which at least 80 percent of the CVAP is of a given population.

Table VI.2. NO-MATCH and MATCH Percent By Racial Group, Using Catalist Racial Classification*

Race	NO-MATCH	MATCH	ALL
Anglo	614,522 (7.4%)	7,649,006 (92.6%)	8,263,528
Black	258,648 (15.1%)	1,453,740 (84.9%)	1,712,388
Hispanic	343,097 (11.3%)	2,704,873 (88.7%)	3,047,970
Other	15,975 (3.3%)	475,810 (96.8%)	491,785
All	1,232,246 (9.1%)	12,283,425 (90.9%)	13,515,671
	Gross Percentage Point Disparity		
Black% – Anglo%	7.7%		
Hispanic% – Anglo%	3.9%		

* Baseline Universe: All Registration Records in TEAM less records indicated as Deceased by State of Texas Database

Table VI.3. NO-MATCH/NOT EXEMPTION ELIGIBLE/NOT AGE VOTE-BY-MAIL ELIGIBLE By Racial Group, Using Catalist Racial Classification*			
RACE	NO MATCH / NOT EXEMPTION ELIGIBLE	NO MATCH / NOT AGE VOTE-BY-MAIL ELIGIBLE	NO-MATCH/ NOT EXEMPTION ELIGIBLE/ NOT AGE VOTE-BY-MAIL ELIGIBLE
Anglo	560,192 (6.8%)	477,010 (5.8%)	436,680 (5.3%)
Black	226,551 (13.2%)	222,570 (13.0%)	196,437 (11.5%)
Hispanic	315,623 (10.4%)	293,214 (9.6%)	272,803 (9.0%)
Other	15,130 (3.1%)	13,215 (2.7%)	12,594 (2.6%)
All	1,117,496 (8.3%)	1,006,009 (7.4%)	918,514 (6.8%)
	Gross Percentage Point Disparity		
Black% – Anglo%	6.4%	7.2%	6.2%
Hispanic% – Anglo%	3.6%	3.8%	3.7%

* Baseline Universe: All Registration Records in TEAM less records indicated as Deceased by State of Texas Database

Table VI.4. Percent NO MATCH and NO MATCH/NOT EXEMPTION ELIGIBLE Among Those Who Voted in 2010 or 2012*				
	NO MATCH		NO MATCH / NOT EXEMPTION ELIGIBLE	
RACE	2010	2012	2010	2012
Anglo	127,274 (3.8%)	228,815 (4.3%)	117,647 (3.5%)	210,644 (4.0%)
Black	48,769 (9.4%)	108, 393 (10.8%)	43,408 (8.3%)	95,905 (9.5%)
Hispanic	38,429 (5.9%)	85,851 (6.4%)	34,495 (5.3%)	77,945 (5.8%)
Other	1,872 (1.8%)	3,894 (1.7%)	1,764 (1.7%)	3,655 (1.6%)
All	216,344 (4.7%)	426,953 (5.4%)	197,314 (4.3%)	388,149 (4.9%)
	Gross Percentage Point Disparity			
Black% – Anglo%	5.6%	6.5%	4.8%	5.5%
Hispanic% – Anglo%	2.1%	2.1%	1.8%	1.8%

* Universe: All Registration Records in TEAM less records indicated as Deceased by State of Texas Database

Table VII.1. Rates of NO MATCHES by Race Under Varying Definitions of the Potential Pool of Registered Voters*				
Race	Excluding Catalist Deceased, Deadwood, or NCOA	Excluding Suspense Voters From Pool of Registered	Excluding Expired ID from Pool of Registered	Excluding Catalist Flagged Records or Suspense Voter or Expired ID
Anglo	544,315 (7.3%)	482,933 (6.6%)	420,733 (5.3%)	317,057 (4.8%)
Black	229,892 (15.0%)	212,078 (14.2%)	188,033 (11.6%)	143,058 (11.0%)
Hispanic	311,227 (11.1%)	291,238 (10.6%)	246,343 (8.5%)	195,190 (8.0%)
Others	14,561 (3.2%)	12,653 (2.8%)	11,501 (2.4%)	8,699 (2.1%)
All	1,099,995 (9.0%)	998,902 (8.3%)	866,610 (6.7%)	664,004 (6.2%)
	Gross Percentage Point Disparity			
Black% – Anglo%	7.7%	7.6%	6.3%	6.2%
Hispanic% – Anglo%	3.8%	4.0%	3.2%	3.2%

* Universe: All Registration Records in TEAM less records indicated as Deceased by State of Texas Database

Table VII.2. Validation of Results With Alternative Racial Classification: Using Records With the Highest Confidence in the Racial Classification*		
Race	NO MATCH	MATCH
Anglo	322,740 (6.7%)	4,501,103 (93.3%)
Black	84,472 (17.5%)	397,065 (82.5%)
Hispanic	138,586 (11.0%)	1,121,722 (89.0%)
Other	5,866 (2.7%)	214,833 (97.3%)
All	551,664 (8.1%)	6,234,719 (91.9%)
	Gross Percentage Point Disparity	
Black% – Anglo%	10.8%	
Hispanic% – Anglo%	4.3%	

*Universe: All Registration Records in TEAM Assign a Racial Classification with High Confidence less records indicated as Deceased by State of Texas Database

Table VII.3. Validation of Results With Alternative Racial Classification Using Spanish Surname Voter Registrations: Comparison of No-Match rates of Spanish Surname Registered Voters and Others*

Race	NO MATCH	MATCH
SSVR	354,652 (11.5%)	2,724,376 (88.5%)
Non-SSVR	877,582 (8.4%)	9,559,052 (91.6%)
All	1,232,234 (9.1%)	12,283,428 (90.9%)
	Gross Percentage Point Disparity	
SSRV – Non-SSRV	3.1%	

* Universe: All Registration Records in TEAM less records indicated as Deceased by State of Texas Database

Table VII.4. Identification Match versus Disability Exemption Eligible*				
		DISABILITY EXEMPTION ELIGIBLE		
STATE OR FEDERAL ID		No Match to Disability	Match to Disability	ALL
	No Match	1,117,496	114,746	1,232,242
	Match	11,450,041	833,388	12,283,429
	ALL	12,567,537	948,134	13,515,671

* Universe: All Registration Records in TEAM less records indicated as Deceased by State of Texas Database

Table VII.5. Identification Database Match Using Department of Justice and State of Texas Algorithms*				
		TEXAS ALGORITHM		
DOJ Algorithm		No Match	Match	ALL
	No Match	1,200,705	31,537	1,232,242
	Match	110,625	12,172,786	12,283,411
	ALL	1,311,330	12,204,323	13,515,653

* Universe: All Registration Records in TEAM less records indicated as Deceased by State of Texas Database

Table VIII.1a. Percent of Registered Anglos, Hispanics, and Blacks in Catalist Database who Voted in the State of Texas in 2010 and 2012: Current Active and Suspense Voters

	2010			2012		
	% Voted	Number Voted	Number Not Voting	% Voted	Number Voted	Number Not Voting
Anglo	41.8%	3,364,053	4,689,493	64.3%	5,169,740	2,874,078
Hispanic	22.0%	655,046	2,320,565	45.0%	1,340,119	1,635,492
Black	31.3%	527,216	1,156,753	59.8%	1,007,153	676,821
	Gross Percentage Point Disparity					
Ang% Vote-Black%Vote	10.5%			4.5%		
Ang% Vote-Hisp% Vote	19.8%			20.3%		

Table VIII.1b. Percent of Registered Anglos, Hispanics, and Blacks in Catalist Database who Voted in the State of Texas in 2010 and 2012: Current Active Voters Only						
	2010			2012		
	% Voted	Number Voted	Number Not Vote	% Voted	Number Voted	Number Not Vote
Anglo	44.0%	3,260,374	4,157,144	68.2%	5,056,818	2,360,700
Hispanic	23.1%	636,741	2,121,984	47.6%	1,312,378	1,446,347
Black	33.2%	509,403	1,024,863	63.4%	973,266	561,000
	Gross Percentage Point Disparity					
Ang% Vote-Black%Vote	9.8%			5.2%		
Ang% Vote-Hisp% Vote	20.9%			20.6%		

Table VIII.2a. Ecological Regression Estimates of Registration as a Percent of Voting Age Population for Anglos, Hispanics, and Blacks in the State of Texas*

(95 Percent Confidence Interval in Parentheses)

	2006	2008	2010	2012
Anglo**	82.5% (± 1.0)	87.1% (± 0.8)	86.0% (± 0.8)	84.7% (± 0.8)
Hispanic	55.0% (± 1.8)	55.0% (± 1.6)	53.0% (± 1.4)	50.5% (± 1.4)
Black	68.5% (± 1.6)	70.5% (± 2.8)	76.6% (± 2.0)	64.7% (± 2.6)
Number of Cases (VTDs)	8660	8660	8660	8660
R-Square	.115	.167	.211	.221
	Gross Percentage Point Disparity			
Anglo%-Black%	14.0%	16.6%	9.4%	20.0%
Anglo%-Hispanic%	27.5%	32.1%	33.0%	34.2%

* Level of analysis: VTD; Dependent variable: Number Registered (on TEAM) divided by CVAP; Weighted by CVAP; Multiple Regression of Percent CVAP Registered on HCVAP Percent and BCVAP Percent.
** Includes Others.

Table VIII.2b. Ecological Regression Estimates of Registration as a Percent of Citizen Voting Age Population for Anglos, Hispanics, and Blacks in the State of Texas*(95 percent confidence intervals in parentheses)

	2006	2008	2010	2012
Anglo**	83.6% (± 0.9)	88.0% (± 0.8)	86.8% (± 0.7)	87.2% (± 0.7)
Hispanic	80.7% (± 2.1)	81.4% (± 1.9)	79.0% (± 1.4)	75.7% (± 1.4)
Black	80.1% (± 3.0)	81.6% (± 2.6)	77.0% (± 2.0)	74.7% (± 2.5)
Number of Cases (VTDs)	8655	8655	8655	8660
R-Square	.001	.007	.007	.026
	Gross Percentage Point Disparity			
Anglo%-Black%	3.4%	6.4%	9.8%	12.6%
Anglo%-Hispanic%	2.9%	6.6%	7.8%	11.6%

* Level of analysis: VTD; Dependent variable: Number Registered (on TEAM) divided by CVAP; Weighted by CVAP; Multiple Regression of Percent CVAP Registered on HCVAP Percent and BCVAP Percent.
** Includes Others.

Table VIII.3a. Ecological Regression Estimates of Voting Rates Among Groups as a Percent of Voting Age Population of Anglos, Hispanics, and Blacks in the State of Texas*

(95 Percent Confidence Interval in Parentheses)

	2006	2008	2010	2012
Anglo**	36.4% (± 0.4)	60.8% (± 0.6)	40.6% (± 0.4)	61.1% (± 0.5)
Hispanic	8.4% (± 0.7)	19.2% (± 0.6)	8.5% (± 0.8)	18.1% (± 1.0)
Black	12.8% (± 1.2)	40.2% (± 1.9)	19.4% (± 1.3)	39.2% (± 1.8)
Number of Cases (VTDs)	8660	8660	8660	8660
R-Square	.398	.392	.442	.431
	Gross Percentage Point Disparity			
Anglo%-Black%	23.6%	20.6%	21.2%	21.8%
Anglo%-Hispanic%	28.0%	41.6%	32.2%	43.0%

* Level of analysis: VTD; Dependent variable: Number Registered (on TEAM) divided by CVAP; Weighted by CVAP; Multiple Regression of Percent CVAP Registered on HCVAP Percent and BCVAP Percent.

** Includes Others.

Table VIII.3b. Ecological Regression Estimates of Voting Rates Among Groups as a Percent of Citizen Voting Age Population of Anglos, Hispanics, and Blacks in the State of Texas*

(95 Percent Confidence Interval in Parentheses)

	2006	2008	2010	2012
Anglo**	36.8% (± 0.4)	61.6% (± 0.5)	41.1% (± 0.4)	61.8% (± 0.5)
Hispanic	13.2% (± 0.9)	29.9% (± 0.6)	13.6% (± 0.9)	28.8% (± 1.2)
Black	15.0% (± 1.2)	44.7% (± 1.9)	21.8% (± 1.3)	43.3% (± 1.8)
Number of Cases (VTDs)	8655	8655	8655	8655
R-Square	.273	.212	.296	.244
	Gross Percentage Point Disparity			
Anglo%-Black%	21.8%	16.9%	19.3%	18.4%
Anglo%-Hispanic%	23.6%	31.7%	27.4%	33.1%

* Level of analysis: VTD; Dependent variable: Number Registered (on TEAM) divided by CVAP; Weighted by CVAP; Multiple Regression of Percent CVAP Registered on HCVAP Percent and BCVAP Percent.

** Includes Others.

Table VIII.4. Current Population Survey Estimates of Percent of Anglo, Hispanic, and Black Adult Citizens who are Registered and who Voted in the State of Texas*				
	Percent Reported Being Registered**			
	2006	2008	2010	2012
Anglo	72.8% (± 3.2)	73.6% (± 3.0)	66.9% (± 3.2)	73.0 % (± 3.0)
Hispanic	58.1% (± 7.0)	54.3% (± 6.0)	53.3% (± 7.2)	54.5% (± 6.8)
Black	64.6% (± 8.8)	74.0% (± 7.2)	61.2% (± 7.8)	72.8 % (± 7.0)
	Percent Reported Voting**			
	2006	2008	2010	2012
Anglo	45.2% (± 3.6)	64.7% (± 3.2)	43.8% (± 3.2)	60.9 % (± 3.2)
Hispanic	25.4% (± 6.2)	37.8% (± 5.8)	23.1% (± 6.0)	38.8% (± 6.6)
Black	36.6% (± 9.0)	65.8% (± 7.8)	37.7% (± 7.8)	62.5% (± 7.6)

*Source: Current Population Survey, various years, "Voting and Registration and Supplement," Table 4b. http://www.census.gov/hhes/www/socdemo/voting/publications/p20/index.htm (last accessed June 6, 2014).
**95 Percent Confidence Interval in Parentheses.

Table VIII.5. Current Population Survey Estimates of Percent of Registered Anglos, Hispanics, and Blacks who Voted in the State of Texas from 2006 to 2012*				
	2006	2008	2010	2012
Anglo	62.5% (± 3.2)	87.8% (± 3.0)	65.4% (± 3.0)	83.4 % (± 3.0)
Hispanic	43.8% (± 7.0)	69.5% (± 6.0)	43.4% (± 7.2)	71.3% (± 6.8)
Black	56.7% (± 8.8)	88.5% (± 7.6)	61.7% (± 7.8)	85.9% (± 7.0)
	Gross Percentage Point Disparity			
Anglo% - Black%	5.8%	-0.7%	3.7%	-2.5%
Anglo% - Hispanic%	18.7%	18.3%	22.0%	12.1%

*Source: Current Population Survey, various years, "Voting and Registration and Supplement," Table 4b. http://www.census.gov/hhes/www/socdemo/voting/publications/p20/index.htm(last accessed June 6, 2014).

APPENDIX

List of Documents Appended Following Appendix Tables

Ex. A – DOJ Algorithm
Ex. B – Texas Algorithm
Ex. C – U.S. State Department Declaration
Ex. D – U.S. Defense Department Declaration
Ex. E – U.S. Citizenship and Immigration Services Declaration
Ex. F– U.S. Social Security Administration Declaration
Ex. G – U.S. Department of Veterans Affairs, Veterans Benefits Administration Declaration
Ex. H – U.S. Department of Veterans Affairs, Veterans Health Administration Declaration

Table A.V.1. Completeness and Uniqueness of Identifiers on TEAM and State of Texas DPS Driver License Databases

Identifier		Texas Registered Voter List (TEAM)		Texas Department of Public Safety (DPS) (DL only)	
		Percent Complete	Percent Unique	Percent Complete	Percent Unique
Primary Matches	A	99.7	100.0	100.0	95.7
	B	99.8	99.5	100.0	95.5
	C	99.8	99.2	100.0	95.4
	D	99.9	99.8	100.0	95.7
	E	99.8	98.0	100.0	94.1
	F	99.8	97.9	100.0	95.4
	M	76.3	98.7	100.0	95.7
Secondary Matches	G	85.0	98.3	88.1	95.9
	H	62.3	99.7	100.0	95.7
	I	62.3	88.7	100.0	95.7
	K	85.0	98.3	88.1	95.9
	L	0.7	100.0	1.1	88.4
	SSN	49.5	98.8	100.0	95.6
Number of Records		13,564,416		16,052,332	

Table A.V.2. Completeness and Uniqueness of Identifiers on State of Texas Voter DPS Public Safety (ID) and Texas License to Carry Databases

Identifier		DPS License to Carry		DPS Personal Identification Card	
		Percent Complete	Percent Unique	Percent Complete	Percent Unique
Primary Matches	A	99.9	100.0	100.0	95.6
	B	99.9	99.5	100.0	96.3
	C	99.9	99.2	100.0	96.2
	D	99.9	99.8	100.0	96.0
	E	100.0	98.0	100.0	96.1
	F	100.0	97.9	100.0	96.4
	M	98.0	98.7	100.0	96.7
Secondary Matches	G	87.6	98.3	79.2	96.6
	H	99.9	99.7	100.0	95.7
	I	99.9	98.0	100.0	95.7
	K	87.6	98.0	79.2	95.9
	L	0.01	98.6	1.7	88.4
	SSN	99.9	98.2	99.5	95.6
Number of Records		835,536		3,396,657	

Table A.V3. Number Matches of TEAM records to State of Texas Identification Databases By Record Identifier				
Identifier		DPS Driver License	DPS Public Safety ID	DPS License to Carry
Primary Matches	A	8,247,986	787,346	533,958
	B	8,422,296	805,092	545,814
	C	8,557,902	829,268	552,372
	D	8,281,572	790,781	535,897
	E	8,195,758	802,261	528,436
	F	9,763,774	1,107,209	676,170
	M	7,953,620	504,694	290,779
Secondary Matches	G	61,458	14,209	5,316
	H	27,408	9,736	2,189
	I	43,291	9,367	3,902
	K	64,523	15,690	5,501
	L	608	226	50
	SSN	107,766	27,451	10,750
Any Identifier		10,663,738	1,284,658	733,008

Table A.V4. Number Matches of TEAM records to Federal Identification Databases By Record Identifier				
Identifier		DOD Military Identification	DOS Passport or Passport Card	USCIS Citizenship Document
Primary Matches	A	350,515	3,018,662	207,387
	B	357,465	3,107,218	221,813
	C	365,985	3,202,237	213,238
	D	352,008	3,036,730	236,996
	E	348,402	3,000,346	206,663
	F	433,861	4,521,683	304,530
Secondary Matches	G	786	11,404	30,225
	H	1,569	31,747	51,524
	I	2,379	23,477	60,063
	K	1,111	16,810	17,378
	L	163	1,001	1,817
	SSN	8,127	150,323	52,938
Any Nationwide Match[2]		166,128	5,386,264	323,425
Any Identifier		638,354	5,731,163	735,086

[2] For DOS, this includes additional matches that were performed in order to deal with a different data storage system used by the Department of State.

Table A.V5. Number Matches of TEAM records to Federal Identification and Exemption Databases By Record Identifier				
Identifier		VA (VHA) VIC/VHIC (Identification)	VA (VBA) Disability (Exemption)	SSA Disability (Exemption)
Primary Matches	A	180,483	100,758	412,295
	B	186,754	103,528	434,720
	C	190,053	105,868	450,148
	D	181,146	101,347	417,348
	E	181,332	101,491	420,879
	F	235,188	146,009	636,806
Secondary Matches	G	454	120,591	2,238
	H	1,063	78,406	9,115
	I	1,509	95,516	5,096
	K	617	120,918	2,704
	L	38	391	236
	SSN	5,951	78,191	268,58
Any Nationwide Match		33,487	25,228	73,236
Any Identifier		296,005	188,516	804,338

Table A.VI.1. Incidence of NO-MATCH By Racial Group, Using Catalist Racial Classification* and Excluding Records indicated as NCOA, Deceased or Deadwood by Catalist			
Race	NO MATCH / NOT EXEMPTION ELIGIBLE	NO MATCH / NOT AGE VOTE-BY-MAIL ELIGIBLE	NO MATCH / NOT EXEMPTION ELIGIBLE/ NOT AGE VOTE-BY-MAIL ELIGIBLE
Anglo	495,081 (6.6%)	429,280 (5.7%)	392,766 (5.3%)
Black	201,036 (13.1%)	199,878 (13.0%)	176,433 (11.5%)
Hispanic	286,269 (10.2%)	269,059 (9.6%)	250,500 (9.0%)
Other	13,788 (3.0%)	12,159 (2.7%)	11,595 (2.6%)
All	996,174 (8.1%)	910,376 (7.4%)	831,294 (6.8%)
	Gross Percentage Point Disparities		
Black % - Anglo %	6.5%	7.3%	6.2%
Hispanic % – Anglo %	3.6%	3.9%	3.7%

Table AVI.3. Estimated Percent No Match by Racial Group Restricting TEAM to ONLY Those With Very High Confidence on Catalist Racial Classification	
Race (Highly Confident)	NO MATCH
Anglo	287,680 (6.6%)
Black	75,714 (17.4%)
Hispanic	125,377 (10.8%)
Other	5,376 (2.6%)
All	494,147 (8.0%)
	Gross Percentage Point Disparity
Black % - Anglo %	10.8%
Hispanic % – Anglo %	4.2%

Exhibit A

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*, Plaintiffs, v. RICK PERRY, *et al.*, Defendants.	Civil Action No. 2:13-cv-193 (NGR)
UNITED STATES OF AMERICA, Plaintiff, TEXAS LEAGUE OF YOUNG VOTERS EDUCATION FUND, *et al.*, Plaintiff-Intervenors, TEXAS ASSOCIATION OF HISPANIC COUNTY JUDGES AND COUNTY COMMISSIONERS, *et al.*, Plaintiff-Intervenors, v. STATE OF TEXAS, *et al.*, Defendants.	Civil Action No. 2:13-cv-263 (NGR)

TEXAS STATE CONFERENCE OF NAACP BRANCHES, *et al.*, Plaintiffs, v. NANDITA BERRY, *et al.*, Defendants.	Civil Action No. 2:13-cv-291 (NGR)
BELINDA ORTIZ, *et al.*, Plaintiffs, v. STATE OF TEXAS, *et al.*, Defendants	Civil Action No. 2:13-cv-348 (NGR)

United States' Database Matching Protocol

The United States hereby sets out its protocol for comparisons among and between the Texas databases and the federal databases. The matching process proceeds in four parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, one-to-many matches are conducted between databases. *Fourth*, Texas data concerning deceased individuals is appended to matching output. Each of the four parts is divided into multiple stages, which in turn are divided into concrete steps.

The database preparation and standardization steps below apply to both the matches requested by the United States and all Plaintiffs, as well as the matches requested by Defendants.

GLOSSARY

The following definitions apply to the terms used in this document.

- **Disability database:** Any federal database containing records that indicate whether an individual has a disability that would permit that individual apply for an exemption from the SB 14 identification requirement.
- **Identification database:** Any Texas or federal database containing records that indicate whether an individual has identification required by SB 14 in most cases to cast an in-person ballot
- **Protocol:** The instructions to prepare data for matching and to conduct multiple matching sweeps between databases according to specified search criteria.
- **Sweep:** A comparison of a set of database fields in the TEAM database against a set of database fields in a disability database or an identification database.
- **TEAM database:** The Texas Election Administration Management database, the state database that contains records of registered voters.

PART I: DATABASE PREPARATION

Stage 1: Extraction of Available Data from TEAM, Identification & Disability Databases

Step 1.1.1: Extract complete name into separate first name, middle name, and last name fields.

Step 1.1.2: Extract date of birth.

Step 1.1.3: Extract gender.

Step 1.1.4: Extract residential address and mailing address.

Step 1.1.5: Extract social security number.

Step 1.1.6: Extract Texas driver license number (only if present in Federal database).

Step 1.1.7: Extract unique record identifier (such as VUID in the TEAM database).

Stage 2: Separate Valid Identification and Disability Records

Step 1.2.1: Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

Step 1.2.2: Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

Step 1.2.3: Separate the identification databases produced by the State of Texas into separate files for each form of state identification at issue, namely driver licenses, identification cards, licenses to carry concealed handguns, and election identification certificates.

Stage 3: Diagnostics

Step 1.3.1: Report the frequency of missing values for each field.

Step 1.3.2: Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.

Step 1.3.3: Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

Stage 4: Standardize Last Name

Step 1.4.1: Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.

Step 1.4.2: For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.

Step 1.4.3: Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.

Step 1.4.4: Code all missing values as blank fields.

Stage 5: Standardize First Name and Middle Name

Step 1.5.1: Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.

Step 1.5.2: Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.[1]

Step 1.5.3: Code all missing values as blank fields.

Stage 6: Standardize Date of Birth

Step 1.6.1: Convert the date of birth to an eight-digit string of MMDDYYYY.

Step 1.6.2: Code all missing values as blank fields.

Stage 7: Standardize Gender

Step 1.7.1: Code gender as a string of 1 for females and 0 for males.

Step 1.7.2: Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.

Step 1.7.3: If missing values remain, code all missing values as blank fields.

[1] The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Stage 8: Standardize Address

Step 1.8.1: Convert the residential ZIP code to a string if it is stored as a numeric field.

Step 1.8.2: Where the residential address ZIP code is blank, populate that field with the value in the mailing address ZIP code field, if available.[2]

Step 1.8.3: Truncate the residential ZIP code field to the first five digits. *E.g.*, <77777-1234> becomes <77777>.

Step 1.8.4: Where the residential address field is blank, populate that field with the value in the mailing address field, if available.

Step 1.8.5: Where address field containing street address begins with a street number, isolate the street number. *E.g.*, <123 Main Street> becomes <123>.

Step 1.8.6: Where the address field begins with recognized strings indicating a mail box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444> becomes <444>.

Step 1.8.7: If missing values remain, code all missing values as blank fields.

Stage 9: Standardize Social Security Number

Step 1.9.1: Convert the social security number to a string if it is stored as a numeric field.

Step 1.9.2: Using full social security number, check for invalid SSNs. In the case of invalid SSNs, code as missing. *E.g.*, <123456789> becomes <>.

[2] For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Step 1.9.3: Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

Step 1.9.4: Code all missing values as blank fields.

Stage 10: Identical Records

Step 1.10.1: For the TEAM database, for the small number of records with different VUID but identical first and last name, gender, residential address number and ZIP, date of birth, and SSN, treat these records as identical.

PART II: DATABASE PREPARATION

Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps

Step 2.1.1: Create Combination A: First Name + Last Name + Gender + Date of Birth + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.[3]

Step 2.1.2: Create Combination B: Last Name + Gender + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.3: Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.4: Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.5: Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.

Step 2.1.6: Create Combination F: First Name + Last Name + Gender + Date of Birth.

Step 2.1.7: Create Combination M: Texas Driver License Number (where available).

[3] For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps

Step 2.2.1: Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.[4]

Step 2.2.2: Create Combination H: SSN4 + Date of Birth + Residential ZIP.

Step 2.2.3: Create Combination I: SSN4 + First Name + Last Name + Date of Birth.

Step 2.2.4: Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.

Step 2.2.5: Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.

Step 2.2.6: Full Social Security Number.

Stage 3: Establish Identifier Uniqueness

Step 2.3.1: Within the TEAM database, determine the uniqueness of each primary and secondary identifier variable and mark accordingly. *E.g.*, if only one record has the string <JEANSMITH01011950012377777> for Combination A, mark that record as unique for Combination A. By contrast, if multiple records have the string <JOHNSMITHA0101950> for Combination G, mark all such records as non-unique for Combination G.

Step 2.3.2: Within the identification and disability databases, generate a field that establishes the uniqueness of each identifier variable. For federal databases, for each combination, generate a field that establishes

[4] Only for the State Department, create three further variations of Combination G using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

uniqueness among only Texas records and a field that establishes uniqueness among nationwide records. *E.g.*, if only one record has the string <JEANSMITH01011950012377777> for Combination A, populate the uniqueness field for Combination A for that record as <1>. If four records have the string <JOHNSMITHA0101950> for Combination G, populate the uniqueness field for Combination G for each of those records as <2>, which indicates any number greater than one.

PART III: MATCH DATABASES

Stage 1: United States' Primary One-to-Many Matching Sweeps

Step 3.1.1: For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identification or disability database. For federal databases, use only the subset of records with Texas addresses in the identification or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identification or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).

Step 3.1.2: Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identification or disability database.

Step 3.1.3: Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identification databases produced by the State of Texas.

Stage 2: United States' Secondary One-to-Many Matching Sweeps

Step 3.2.1: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identification or disability database. For federal databases, use only the subset of records with Texas addresses in the identification or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identification or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.2.2: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number[5] in the TEAM database against the equivalent combination/field in the identification or disability database.

Step 3.2.3: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identification or disability database.

[5] The full social security number is not created as a separate "combination" as it is its own field stored within the TEAM database under the field name "ssn".

Step 3.2.4: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identification or disability database.

Stage 3: United States' Nationwide Federal Database Sweeps

Step 3.3.1: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identification or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identification or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identification or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.3.2: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identification or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identification or disability database.

Step 3.3.3: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identification or disability database, use the procedure in Step 3.3.1 to match Combination

K against Combination G, Combination K, and Combination L in the nationwide identification or disability database.

Step 3.3.4: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identification or disability database, use the procedure in Step 3.3.1 to match Combination L against Combination G, Combination K, and Combination L in the nationwide identification or disability database.[6]

[6] Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last_fix + First_fix + Middle_name from the TEAM database;
- Combination G2 to DOB + Last_fix + First_fix from the TEAM database;
- Combination G2 to DOB + Last_fix + First word of First_name from the TEAM database;
- Combination G3 to DOB + Last_fix + First_fix + Middle_Initial from the TEAM database;
- Combination G3 to DOB + Last_fix + First word of First_name + Middle_Initial from the TEAM database; and
- Combination G3 to DOB + Last_fix + First word of First_name + First character of Second word of First_fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

PART IV: DATA GATHERING

Stage 1: Gather Information Regarding Deceased Individuals

Step 4.1.1: In all instances in which a unique match was achieved between the TEAM database and the Texas Driver License, Texas Identification Card, and Texas Concealed Handgun database, determine whether the identification record at issue has been flagged as deceased in the Texas identification database.

Step 4.1.2: Where the driver license, concealed handgun license, or Texas identification card record at issue has been flagged as deceased, append the "deceased" flag to the TEAM record to which the unique match has been made.

Criteria for Matching Records in the TEAM Database

ID Matching			
Database	**Valid ID**	**Matching Algorithm(s)**	**Additional Parameters**
Texas Dept. of Public Safety	Driver's License	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB 5. DL Number	ID unexpired or expired within 60 days of date that TEAM database was extracted
	Personal Identification Card	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB 5. DL Number	ID unexpired or expired within 60 days of date that TEAM database was extracted
	Concealed Handgun License	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB 5. DL Number	ID unexpired or expired within 60 days of date that TEAM database was extracted
	Texas Election Identification Certificate	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB 5. DL Number	ID unexpired or expired within 60 days of date that TEAM database was extracted
U.S. Department of Defense	Military ID-Common Access Card	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	ID unexpired or expired within 60 days of date that TEAM database was extracted
	Military ID- Uniformed Services ID Cards	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	ID unexpired or expired within 60 days of date that TEAM database was extracted
	Military ID- DOD Civilian Retiree Card	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	ID unexpired or expired within 60 days of date that TEAM database was extracted

U.S. Dept. of Veterans Affairs	Veterans Affairs ID Card	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	None.
U.S. Dept. of Homeland Security	U.S. Certificate of Citizenship	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	None
	U.S. Certificate of Naturalization	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	None
U.S. State Department	U.S. Passport	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	ID unexpired or expired within 60 days of date that TEAM database was extracted
	U.S. Passport Card	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	ID unexpired or expired within 60 days of date that TEAM database was extracted
Permanent Exemptions			
U.S. Social Security Administration	N/A	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	Determined to have a disability
U.S. Department of Veterans Affairs	N/A	1. SSN (Last Four) + Last Name + DOB 2. SSN (Full) 3. First Name +Last Name + DOB 4. First Name+Middle Initial+Last Name+DOB	Disability rating of 50% or more

Notes:
For each ID/Database combination outlined above all specified matching queries should be run. Based on the matching algorithms, if a record (registrant) in the TEAM database matches a record in the database being searched, then this will be denoted as a match (For any given type of SB14 ID, a registrant will be determined to have valid ID if matched by any of the matching algorithms outlined). New fields will be added to the TEAM database to denote the presence of a match for each database/ID/query. In addition to searching for ID possession, records in the TEAM database will be flagged to indicate the registrant is exempt due to the presence of a qualified disability. (Disclaimer: The matching algorithms here denoted are based on data fields currently known to populate the listed databases).

Explanation of matching queries to be run (numbers correspond to matching algorithms in table above):
1. Records will be matched using the last four digits of the Social Security Number (where available) along with the last name and date of birth fields.
2. Records will be matched using the full Social Security Number field (where available).
3. Records will be matched using the first name, last name, and date of birth fields.
4. Records will be matched using the first name, middle initial, last name, and date of birth fields.
5. Records will be matched using the Texas Driver's License Number field (where available).

Exhibit C

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*, Plaintiffs, v. RICK PERRY, *et al.*, Defendants.	Civil Action No. 2:13-cv-193 (NGR)
UNITED STATES OF AMERICA, Plaintiff, TEXAS LEAGUE OF YOUNG VOTERS EDUCATION FUND, *et al.*, Plaintiff-Intervenors, TEXAS ASSOCIATION OF HISPANIC COUNTY JUDGES AND COUNTY COMMISSIONERS, *et al.*, Plaintiff-Intervenors, v. STATE OF TEXAS, *et al.*, Defendants.	Civil Action No. 2:13-cv-263 (NGR)

TEXAS STATE CONFERENCE OF NAACP BRANCHES, *et al.*, Plaintiffs, v. JOHN STEEN, *et al.*, Defendants.	Civil Action No. 2:13-cv-291 (NGR)
BELINDA ORTIZ, *et al.*, Plaintiffs, v. STATE OF TEXAS, *et al.*, Defendants	Civil Action No. 2:13-cv-348 (NGR)

CORRECTED DECLARATION OF LEE CHARLES BAYDUSH

I, Lee Charles Baydush, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as subject matter expert for the U.S. Department of State, Bureau of Consular Affairs, Consular Systems & Technology, Service Integration & Innovation Division (CA/CST/SII). I have served Consular Affairs in this capacity with SII and others as a contractor specializing in Passport Systems for more than 28 years. My duties include research, analysis, consulting, and operational support of Consular IT systems.
2. I have personal knowledge of the information contained in this declaration based upon my work for the State Department, and my personal involvement in completion of the database comparison described below.

3. On February 27, 2014, the State Department received an encrypted hard drive containing data regarding Texas registered voters from the United States Department of Justice in the form of a 7 gigabyte .csv file containing 13,564,420 records for Texas registered voters. The State Department was able to load all records except for 6. The identifiers (VUIDs) for 5 of the 6 records that my agency was not able to load and analyze as part of the database comparison process are as follows: 1018504714; 1018510162; 1018555565; 1111894862; 1164253546. The final record that could not be loaded lacked a VUID, but has a unique identifier of 3890395vf (other than this unique identifier "var_fill" value, this record was blank).

4. The Texas data that the State Department received contained the following fields:

- Vuid
- last_name
- first_name
- middle_name
- ssn
- ssn4_str
- gender_code
- official_id
- dl_id_str
- date_of_birth
- voter_status_code
- block_number
- street_name
- city
- state
- zip_code
- zip_str
- mail_adrs_1
- last_fix
- multi_lastname1
- multi_lastname2
- first_fix
- middle_initial
- fem_str
- dob_str
- zip5_res_alt
- str_num_res_alt
- AGDN
- AGDN_TAG
- AGDN_unique
- AGDNlast
- AGDNlast_TAG
- AGDNlast_unique
- AGD
- AGD_TAG
- AGD_unique
- and
- ADN_TAG
- ADN_unique
- AGN
- AGN_TAG
- AGN_unique
- GDN
- GDN_TAG
- GDN_unique
- DLN
- DLN_TAG
- DLN_unique

- unique_count
- FMLD
- FMLD_TAG
- FMLD_unique
- SDZ
- SDZ_TAG
- SDZ_unique
- SND
- SND_TAG
- SND_unique
- FML1D
- FML1D_TAG
- FML1D_unique
- FML2D
- FML2D_TAG
- FML2D_unique
- unique_count2
- var_fill
- K2
- K3
- L2
- L3

5. The State Department executed the data preparation and comparison steps, attached as Ex. A, that were provided by the United States Department of Justice.

6. Through this process, the State Department attempted to match particular combinations of identifying information for Texas registered voters (*e.g.*, first and last name, gender, and date of birth) with the same or related combinations of identifying information with respect to holders of U.S. Passports and U.S. Passport cards. For each particular combination, the State Department identified instances where the identifying information for a Texas registered voter matched with the respective combination of identifying information for one or more holders of a U.S. Passport or U.S. Passport Card.

7. For each sweep through the relevant State Department datasets, we appended a column to the Texas data to indicate, on a record-by-record basis, the output of the database comparisons on all of the requested combinations.

8. Two State Department datasets were created for the analysis—one contained qualifying records of U.S. Passport and U.S. Passport Card holders with a Texas mailing address and the other contained qualifying records of U.S. Passport and U.S. Passport Card holders without a Texas mailing address. The frequencies of any missing values in those datasets are provided as Ex. B.

9. On April 9, 2014, the State Department transferred responsive data to the Department of Justice, in the form of a variable width file with comma delimiters. In this file, there were 7,741 records in the TEAM database with commas embedded in the address field, and several hundred records with commas embedded in various name fields. Such commas could be incorrectly treated as delimiters in reading the Department of State matching results. As a result, I re-exported the results file using pipe delimiters and on May 14, 2014, the State Department transferred to the Department of Justice a compressed and encrypted 1.7 gigabyte zip file on a DVD that contained a 7.44 gigabyte pipe-delimited file. Both the April 9 and May 14 files are identical other than the use of different delimiters. The use of pipe delimiters on the May 14, 2014, file will prevent any difficulties reading the embedded commas in the results file. The name of the May 14, 2014 results file is Team_Pipe_Delim.
10. The results of each sweep, as described in Ex. A, can be found in the following columns of the data that the State Department returned to the Department of Justice on May 14:
    a. "AGDN_matches_id" contains results of the Stage 1, Step 3.1.1, Combination A match;
    b. "AGDNlast_matches_id" contains results of the Stage 1, Step 3.1.2, Combination B match;
    c. "AGD_matches_id" contains results of the Stage 1, Step 3.1.2, Combination C match;
    d. "ADN_matches_id" contains results of the Stage 1, Step 3.1.2, Combination D match;
    e. "AGN_matches_id" contains results of the Stage 1, Step 3.1.2, Combination E match;
    f. "GDN_matches_id" contains results of the Stage 1, Step 3.1.2, Combination F match;

g. "FMLD_matches_id" contains results of the Stage 2, Step 3.2.1, Combination G match;

h. "SDZ_matches_id" contains results of the Stage 2, Step 3.2.2, Combination H match;

i. "SND_matches_id" contains results of the Stage 2, Step 3.2.2, Combination I match;

j. "ssn_matches_id" contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;

k. "k2_matches_id" contains results of the Stage 2, Step 3.2.3, Combination K to G match;

l. "FML1D_matches_id" contains results of the Stage 2, Step 3.2.3, Combination K match;

m. "k3_matches_id" contains results of the Stage 2, Step 3.2.3, Combination K to L match;

n. "l2_matches_id" contains results of the Stage 2, Step 3.2.4, Combination L to G match;

o. "FML2D_matches_id" contains results of the Stage 2, Step 3.2.4, Combination L match;

p. " l3_matches_id" contains results of the Stage 2, Step 3.2.4, Combination L to K match;

q. "GDN_matches_id_Other_GDN" contains results of the Stage 3, Step 3.3.1, Combination F match;

r. "FMLD_matches_id_Other_FMLD" contains results of the Stage 3, Step 3.3.2, Combination G match;

s. "SND_matches_id_Other_SND" contains results of the Stage 3, Step 3.3.2, Combination I match;

t. "ssn_matches_id_Other" contains results of the Stage 3, Step 3.3.2, nine-digit social security number match;

u. "FML1D_matches_id_Other_FMLD" contains results of the Stage 3, Step 3.3.3, Combination K to G match;

v. "FML1D_matches_id_Other_FML1D" contains results of the Stage 2, Step 3.3.3, Combination K match;

w. "FML1D_matches_id_Other_FML2D" contains results of the Stage 2, Step 3.3.3, Combination K to L match;

x. "FML2D_matches_id_Other_FMLD" contains results of the Stage 3, Step 3.3.4, Combination L to G match;

y. "FML2D_matches_id_Other_FML2D" contains results of the Stage 3, Step 3.3.4, Combination L match;

z. "FML2D_matches_id_Other_FML1D" contains results of the Stage 3, Step 3.3.4, Combination L to K match;

aa. "a1_matches_id_TX" and "a1_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G1 to DOB + Last_fix + First_fix + Middle_name matches for the Texas-only and non-Texas address datasets;

bb. "a2_1_matches_id_TX" and "a2_1_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G2 to DOB + Last_fix + First_fix matches for the Texas-only and non-Texas address datasets;

cc. "a2_2_matches_id_TX" and "a2_2_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G2 to DOB + Last_fix + First word of First_name matches for the Texas-only and non-Texas address datasets;

dd. "a3_1_matches_id_TX" and "a3_1_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G3 to DOB + Last_fix + First_fix + Middle_Initial matches for the Texas-only and non-Texas address datasets;

ee. "a3_2_matches_id_TX" and "a3_2_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G3 to DOB Last_fix + First word of First_name + Middle_Initial matches for the Texas-only and non-Texas address datasets;

ff. "a3_3_matches_id_TX" and "a3_3_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G3 to DOB Last_fix + First word of First_name + First character of Second word of First_fix matches for the Texas-only and non-Texas address datasets;

gg. "TXMM_snlastdAll_matches_id" contains results of the Stage 4, Step 3.4.1, Sweep 1 match, while "TXMM_TX_snlastdAll_matches_id" contains the results for the Sweep 1 match based on the Texas-only dataset and "TXMM_Other_snlastdAll_matches_id" contains the results for the Sweep 1 match based on the non-Texas addresses dataset;

hh. "TXMM_ssn_matches_id" contains results of the Stage 4, Step 3.4.2, Sweep 2 match, while "TXMM_TX_ssn_matches_id" contains the results for the Sweep 2 match based on the Texas-only dataset and "TXMM_Other_ssn_matches_id" contains the results for the Sweep 2 match based on the non-Texas address dataset;

ii. "TXMM_ndAll_matches_id" contains results of the Stage 4, Step 3.4.3, Sweep 3 match, while "TXMM_TX_ndAll_matches_id" contains the results for the Sweep 3 match based on the Texas-only dataset and "TXMM_Other_ndAll_matches_id" contains the results for the Sweep 3 match based on the non-Texas address dataset; and

jj. "TXMM_fmldAll_matches_id" contains results of the Stage 4, Step 3.4.4, Sweep 4 match, while "TXMM_TX_fmldAll_matches_id" contains the results for the Sweep 4 match based on the Texas-only dataset and "TXMM_Other_fmldAll_matches_id" contains the results for the Sweep 4 match based on the non-Texas address dataset.

11. No other matching comparisons between the Texas data and data from State Department were undertaken beyond those set forth above, in accordance with the steps provided in Ex. A.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on June 24, 2014.

Lu Baydush

Exhibit A

***United States v. Texas*: Federal Agency Algorithm Summary**

This document summarizes the database matches that both the United States and all other plaintiffs and the State of Texas have requested from Federal Agencies as part of the *Veasey v. Perry*/*United States v. Texas* litigation (S.D. Tex).

The matching process proceeds in three parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, the United States' one-to-many matches and the State of Texas's many-to-many matches are conducted between databases.

PART I: DATABASE PREPARATION

Stage 1: Extraction of Available Data from Federal Identification & Disability Databases

Step 1.1.1: Extract complete name into separate first name, middle name, and last name fields.

Step 1.1.2: Extract date of birth.

Step 1.1.3: Extract gender.

Step 1.1.4: Extract residential address and mailing address.

Step 1.1.5: Extract social security number.

Step 1.1.6: Extract Texas driver license number (only if present in Federal database).

Stage 2: Separate Valid Identification and Disability Records

Step 1.2.1: Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

Step 1.2.2: Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

Stage 3: Diagnostics

Step 1.3.1: Report the frequency of missing values for each field.

Step 1.3.2: Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.

Step 1.3.3: Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

Stage 4: Standardize Last Name

Step 1.4.1: Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.

Step 1.4.2: For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.

Step 1.4.3: Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.

Step 1.4.4: Code all missing values as blank fields.

Exhibit A

Stage 5: Standardize First Name and Middle Name

Step 1.5.1: Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.

Step 1.5.2: Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.[1]

Step 1.5.3: Code all missing values as blank fields.

Stage 6: Standardize Date of Birth

Step 1.6.1: Convert the date of birth to an eight-digit string of MMDDYYYY.

Step 1.6.2: Code all missing values as blank fields.

Stage 7: Standardize Gender

Step 1.7.1: Code gender as a string of 1 for females and 0 for males.

Step 1.7.2: Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.

Step 1.7.3: If missing values remain, code all missing values as blank fields.

[1] The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Stage 8: Standardize Address

Step 1.8.1: Convert the residential ZIP code to a string if it is stored as a numeric field.

Step 1.8.2: Where the residential address ZIP code is blank, populate that field with the value in the mailing address ZIP code field, if available.[2]

Step 1.8.3: Truncate the residential ZIP code field to the first five digits. *E.g.*, <77777-1234> becomes <77777>.

Step 1.8.4: Where the residential address field is blank, populate that field with the value in the mailing address field, if available.

Step 1.8.5: Where address field containing street address begins with a street number, isolate the street number. *E.g.*, <123 Main Street> becomes <123>.

Step 1.8.6: Where the address field begins with recognized strings indicating a mail box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444> becomes <444>.

Step 1.8.7: If missing values remain, code all missing values as blank fields.

Stage 9: Standardize Social Security Number

Step 1.9.1: Convert the social security number to a string if it is stored as a numeric field.

Step 1.9.2: Using full social security number, check for invalid SSNs. In the case of invalid SSNs, code as missing. *E.g.*, <123456789> becomes <>.

[2] For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Step 1.9.3: Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

Step 1.9.4: Code all missing values as blank fields.

PART II: DATABASE PREPARATION

Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps

Step 2.1.1: Create Combination A: First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.[3]

Step 2.1.2: Create Combination B: Last Name + Gender + DOB + Residential ZIP + Residential Street Number.

Step 2.1.3: Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.4: Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.5: Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.

Step 2.1.6: Create Combination F: First Name + Last Name + Gender + DOB.

Step 2.1.7: Create Combination M: Texas Driver License Number (where available).

[3] For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Exhibit A

Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps

Step 2.2.1: Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.[4]

Step 2.2.2: Create Combination H: SSN4 + Date of Birth + Residential ZIP.

Step 2.2.3: Create Combination I: SSN4 + First Name + Last Name + Date of Birth.

Step 2.2.4: Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.

Step 2.2.5: Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.

Step 2.2.6: Full Social Security Number.

Stage 3: Construct Identifiers Used Only For Texas's Many-to-Many Sweeps

Step 2.3.1: Create Combination for Texas's Sweep 1: SSN4 + Last Name + DOB.

Step 2.3.2: Create Combination for Texas's Sweep 3: First Name + Last Name + DOB

Step 2.3.3: Create Combination for Texas's Sweep 4: First Name + Middle Initial + Last Name + DOB

Note: Combinations for Texas's Sweeps 1 and 3 do not already exist as pre-made fields in the TEAM database extract but instead must be created from the underlying TEAM database fields, in addition to being constructed on the Federal database side. Texas's Sweep 4 is equivalent to

[4] Only for the State Department, create three further variations of Combination G created using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

the combination for the United States' Combination G. Texas's Sweep 2 is on full 9 social security number.[5]

Stage 4: Establish Identifier Uniqueness For Combinations A - L

Step 2.4.1: Generate a field that establishes the uniqueness of each identifier variable. For federal databases, for each combination A-L, generate a field that establishes uniqueness among only Texas records and a field that establishes uniqueness among nationwide records. *E.g.*, if only one record has the string <JEANSMITH01011950012377777> for Combination A, populate the uniqueness field for Combination A for that record as <1>. If four records have the string <JOHNSMITHA0101950> for Combination G, populate the uniqueness field for Combination G for each of those records as <2>, which indicates any number greater than one.

[5] For purposes of matching to U.S. Department of State Passport and Passport Card holder data, Texas's Sweep 3 and Sweep 4 are as follows: Sweep 3: Last Name + First Name (restricted to 32 characters) + DOB. Sweep 4: Last Name + First Name + Middle Initial (restricted to 32 characters) + DOB.

Exhibit A

PART III: MATCH DATABASES

Stage 1: United States' Primary One-to-Many Matching Sweeps

Step 3.1.1: For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identifier or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).

Step 3.1.2: Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identifier or disability database.

Step 3.1.3: Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identifier databases produced by the State of Texas.

Exhibit A

Stage 2: United States' Secondary One-to-Many Matching Sweeps

Step 3.2.1: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identifier or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.2.2: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number[6] in the TEAM database against the equivalent combination/field in the identifier or disability database.

Step 3.2.3: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identifier or disability database.

[6] The full social security number is not created as a separate "combination" as it is its own field stored within the TEAM database under the field name "ssn".

Exhibit A

Step 3.2.4: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identifier or disability database.

Stage 3: United States' Nationwide Federal Sweeps

Step 3.3.1: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identifier or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.3.2: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identifier or disability database.

Step 3.3.3: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination

K against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.

Step 3.3.4: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination L against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.[7]

Stage 4: Texas' Many-to-Many Nationwide Sweeps

Step 3.4.1 Regardless of whether the combination for Sweep 1 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

[7] Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last_fix + First_fix + Middle_name from the TEAM database;
- Combination G2 to DOB + Last_fix + First_fix from the TEAM database;
- Combination G2 to DOB + Last_fix + First word of First_name from the TEAM database;
- Combination G3 to DOB + Last_fix + First_fix + Middle_Initial from the TEAM database;
- Combination G3 to DOB + Last_fix + First word of First_name + Middle_Initial from the TEAM database; and
- Combination G3 to DOB + Last_fix + First word of First_name + First character of Second word of First_fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

Step 3.4.2 Regardless of whether full 9 social security number is unique in the TEAM database, for Sweep 2, match against the equivalent field in a nationwide search of the Federal database.

Step 3.4.3 Regardless of whether the combination for Sweep 3 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Step 3.4.4 Regardless of whether the combination for Sweep 4 (Combination G) is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Note: For each of the Texas many-to-many sweeps:

- Indicate <1> if any TEAM combination matches a single combination in the Federal database
- Indicate <2> if any TEAM combination matches more than one record in the Federal database.
- Indicate <0> if no match is achieved.

Examples:

- If there are two TEAM records that have identical versions of the combination for Sweep 1, and there is one record in the Federal database that matches on this combination, both of the underlying TEAM records will have a matching output of <1> for Sweep 1.

- If there are three TEAM records that have identical versions of the combination for Sweep 3, and there are five records in the Federal database that match on that combination, the three TEAM records will each have a matching output of <2> for Sweep 3.

Exhibit B

Step 1.3.1 Diagnostics: Report the Frequency of Missing Values for Each Field

6, 781,544 Qualifying U.S. Passport/U.S. Passport Card records with 'TX' as the mailing address state

Field Name	Records with Missing Values
last_name	0
first_name	38
date_of_birth	0
gender_code	5
ssn	25,640
Last_fix	0
multi_lastname1	6,717,750
multi_lastname2	6,717,749
first_fix	38
middle_initial	929,454
fem_str	0
dob_str	0
zip5_num_res_alt	50
ssn4_str	25,640

92,776,332 Qualifying U.S. Passport/Passport Card records without 'TX' as the mailing address state

Field Name	Records with Missing Values
last_name	0
first_name	1,596
date_of_birth	2
gender_code	247
ssn	641,147
last_fix	0
multi_lastname1	91,615,428
multi_lastname2	91,615,406
first_fix	1,596
middle_initial	12,567,280
fem_str	16
dob_str	0
zip5_res_alt	757,630
str_num_res_alt	43,961

Exhibit D

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*, Plaintiffs, v. RICK PERRY, *et al.*, Defendants.	Civil Action No. 2:13-cv-193 (NGR)
UNITED STATES OF AMERICA, Plaintiff, TEXAS LEAGUE OF YOUNG VOTERS EDUCATION FUND, *et al.*, Plaintiff-Intervenors, TEXAS ASSOCIATION OF HISPANIC COUNTY JUDGES AND COUNTY COMMISSIONERS, *et al.*, Plaintiff-Intervenors, v. STATE OF TEXAS, *et al.*, Defendants.	Civil Action No. 2:13-cv-263 (NGR)

TEXAS STATE CONFERENCE OF NAACP BRANCHES, *et al.*, Plaintiffs, v. JOHN STEEN, *et al.*, Defendants.	Civil Action No. 2:13-cv-291 (NGR)
BELINDA ORTIZ, *et al.*, Plaintiffs, v. STATE OF TEXAS, *et al.*, Defendants	Civil Action No. 2:13-cv-348 (NGR)

DECLARATION OF Michelle Saunders Rudolph

I, Michelle Saunders Rudolph, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as Acting Division Director, Data, Analysis and Programs Division, DMDC/DHRA/DoD. I have served in this capacity for close to 2 years. My duties include ensuring the data arriving at DMDC is of sound quality so Servicemembers and their families can receive the benefits and entitlements they have earned. In addition, I ensure agencies and offices with a need to know, receive accurate and timely data.
2. I have personal knowledge of the information contained in this declaration based upon my work for the Defense Manpower Data Center at the U.S. Department of Defense ("DMDC"), and my supervision of completion of the database comparison described below.

3. On February 28, 2014, DMDC received from the United States Department of Justice an encrypted hard drive containing data regarding Texas registered voters in the form of a 7 gigabyte .csv file containing 13,564,420 records for Texas registered voters. DMDC was able to load all 13,564,420 records.

4. The Texas data that DMDC received contained the following fields:

- Vuid
- last_name
- first_name
- middle_name
- ssn
- ssn4_str
- gender_code
- official_id
- dl_id_str
- date_of_birth
- voter_status_code
- block_number
- street_name
- city
- state
- zip_code
- zip_str
- mail_adrs_1
- last_fix
- multi_lastname1
- multi_lastname2
- first_fix
- middle_initial
- fem_str
- dob_str
- zip5_res_alt
- str_num_res_alt
- AGDN
- AGDN_TAG
- AGDN_unique
- AGDNlast
- AGDNlast_TAG
- AGDNlast_unique
- AGD
- AGD_TAG
- AGD_unique
- and
- ADN_TAG
- ADN_unique
- AGN
- AGN_TAG
- AGN_unique
- GDN
- GDN_TAG
- GDN_unique
- DLN
- DLN_TAG
- DLN_unique
- unique_count
- FMLD
- FMLD_TAG
- FMLD_unique
- SDZ
- SDZ_TAG
- SDZ_unique
- SND
- SND_TAG
- SND_unique
- FML1D
- FML1D_TAG
- FML1D_unique
- FML2D
- FML2D_TAG
- FML2D_unique
- unique_count2
- var_fill

- K2
- K3
- L2
- L3

5. DMDC executed the data preparation and comparison steps, attached as Ex. A, that were provided by the United States Department of Justice.

6. Through this process, DMDC attempted to match particular combinations of identifying information for Texas registered voters (*e.g.*, first and last name, gender, and date of birth) with the same or related combinations of identifying information with respect to persons eighteen years of age or older holding a valid form of one of the following Department of Defense ("DoD") photo identification cards: Common Access Cards and Uniformed Services ID Cards (including Department of Defense Civilian Retiree Cards). For each particular combination, DMDC identified instances where the identifying information for a Texas registered voter matched with the respective combination of identifying information for one or more holders of DoD photo identification.

7. In order to conduct the requested matches, and to create and record the combinations used for the Stage 4 matches, DMDC reformatted the original .csv file into a fixed format flat file to which matching results were appended.

8. Two DoD datasets were created for this analysis—one contained records of holders of qualifying DoD photo ID located in Texas and the other contained records for holders of qualifying DoD photo ID located outside of Texas. For each sweep through the relevant DoD datasets, we appended a column to the Texas data to indicate, on a record-by-record basis, the output of the database comparisons on all of the requested combinations. The frequencies of any missing values in the underlying DoD data are provided as Ex. B.

9. On May 9, 2014, DMDC transferred responsive data to the Department of Justice, in the form of a 21.5 gigabyte file with pipe delimiters. The name of the results file transferred to the Department of Justice is "Team Match Report".

10. The results of each sweep, as described in Ex. A, can be found in the following columns of the data that DMDC returned to the Department of Justice on May 9:

    a. "Key_A_match_flag_Texas" contains results of the Stage 1, Step 3.1.1, Combination A match;

    b. "Key_B_match_flag_Texas" contains results of the Stage 1, Step 3.1.2, Combination B match;

    c. "Key_C_match_flag_Texas" contains results of the Stage 1, Step 3.1.2, Combination C match;

    d. "Key_D_match_flag_Texas" contains results of the Stage 1, Step 3.1.2, Combination D match;

    e. "Key_E_match_flag_Texas" contains results of the Stage 1, Step 3.1.2, Combination E match;

    f. "Key_F_match_flag_Texas" contains results of the Stage 1, Step 3.1.2, Combination F match;

    g. "Key_G_match_flag_Texas" contains results of the Stage 2, Step 3.2.1, Combination G match;

    h. "Key_H_match_flag_Texas" contains results of the Stage 2, Step 3.2.2, Combination H match;

    i. "Key_I_match_flag_Texas" contains results of the Stage 2, Step 3.2.2, Combination I match;

j. "Key_KtoG_match_flag_Texas" contains results of the Stage 2, Step 3.2.3, Combination K to G match;

k. "Key_KtoK_match_flag_Texas" contains results of the Stage 2, Step 3.2.3, Combination K match;

l. "Key_KtoL_match_flag_Texas" contains results of the Stage 2, Step 3.2.3, Combination K to L match;

m. "Key_LtoG_match_flag_Texas" contains results of the Stage 2, Step 3.2.4, Combination L to G match;

n. "Key_LtoL_match_flag_Texas" contains results of the Stage 2, Step 3.2.4, Combination L match;

o. "Key_LtoK_match_flag_Texas" contains results of the Stage 2, Step 3.2.4, Combination L to K match;

p. "Key_F_match_flag_Nationwide" contains results of the Stage 3, Step 3.3.1, Combination F match;

q. "Key_I_match_flag_Nationwide" contains results of the Stage 3, Step 3.3.2, Combination I match;

r. "Key_KtoG_match_flag_Nationwide" contains results of the Stage 3, Step 3.3.3, Combination K to G match;

s. "Key_KtoK_match_flag_Nationwide" contains results of the Stage 2, Step 3.3.3, Combination K match;

t. "Key_KtoL_match_flag_Nationwide" contains results of the Stage 2, Step 3.3.3, Combination K to L match;

u. "Key_LtoG_match_flag_Nationwide" contains results of the Stage 3, Step 3.3.4, Combination L to G match;

v. "Key_LtoL_match_flag_Nationwide" contains results of the Stage 3, Step 3.3.4, Combination L match;

w. "Key_LtoK_match_flag_Nationwide" contains results of the Stage 3, Step 3.3.4, Combination L to K match;

11. In the May 9 dataset provided to the Department of Justice, no results were provided for the following matches:

a. the Stage 2, Step 3.2.2, nine-digit social security number match;

b. the Stage 3, Step 3.3.2, Combination G match; and

c. the Stage 3, Step 3.3.2, nine-digit social security number match.

These match results were provided separately on to the Department of Justice on May 28, 2014 as described in paragraph 13 below.

12. In the original May 9 dataset provided to the Department of Justice, results for the Stage 4 matches were provided in the following fields:

a. "Sweep_1_match_flag_Nationwide" had partial results for the Stage 4, Step 3.4.1, Sweep 1 match;

b. "Sweep_2_match_flag_Nationwide" had partial results for the Stage 4, Step 3.4.2, Sweep 2 match;

c. "Sweep_3_match_flag_Nationwide" had partial results for the Sweep 3 match; and

d. "Sweep_4_match_flag_Nationwide" had partial results for the Stage 4, Step 3.4.4, Sweep 4 match.

For these Stage 4 matches in the May 9 dataset, the only voter registration records compared were those that had not already matched on a Stage 1 through 3 sweep. Moreover, the only DoD records compared were for qualifying ID holders outside of Texas. At the request of the Department of Justice, DMDC re-ran each of the Stage 4 matches, matching all Texas voter registration records regardless of whether they had been matched on a Stage 1 through 3 sweep. In addition, the complete voter registration database extract was matched against all qualifying DoD records for ID holders in and outside of Texas. These corrected and complete match results for the Stage 4 sweeps were provided to Department of Justice as part of the May 28, 2014 dataset.

13. The May 28 dataset thus supplemented data not previously provided by DMDC as to three of the sweeps requested by the United States, and also re-ran the each of the Stage 4 matches requested by Defendants. To enable the results from the May 28 dataset to be integrated with the May 9 dataset, the May 28 dataset included two unique identifiers: the VUID field from the voter registration file, which is almost completely unique, as well as the Var_Fill field, which is completely unique. In addition to these identifiers, the following results fields were included in the May 28 dataset:

    a. "SWEEP_1_MATCH_FLAG_TEXAS" contains results of the Stage 4, Step 3.4.1, Sweep 1 match based on the Texas-only dataset and "SWEEP_1_MATCH_FLAG_NATIONWIDE" contains the results for the same step based on the non-Texas, nationwide dataset;

    b. "SWEEP_2_MATCH_FLAG_TEXAS" contains results of the Stage 4, Step 3.4.2, Sweep 2 match based on the Texas-only dataset and

"SWEEP_2_MATCH_FLAG_NATIONWIDE" contains the results for the same step based on the non-Texas, nationwide dataset;

c. "SWEEP_3_MATCH_FLAG_TEXAS" contains results of the Stage 4, Step 3.4.3, Sweep 3 match based on the Texas-only dataset and "SWEEP_3_MATCH_FLAG_NATIONWIDE" contains the results for the same step based on the non-Texas, nationwide dataset;

d. ""SWEEP_4_MATCH_FLAG_TEXAS" contains results of the Stage 4, Step 3.4.4, Sweep 4 match based on the Texas-only dataset and "SWEEP_4_MATCH_FLAG_NATIONWIDE" contains the results for the same step based on the non-Texas, nationwide dataset;

e. "SSN9_SECONDARY_TEXAS" contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;

f. "SSN9_SECONDARY_NATIONWIDE" contains the results of the Stage 3, Step 3.3.2, nine-digit social security number match; and

g. "KEY_G_MATCH_FLAG_NATIONWIDE" contains the results of the Stage 3, Step 3.3.2, Combination G match.

14. In both the May 9 and May 28 datasets, a result of "1" indicates that the respective combination from one DoD record matched; "2" indicates that two or more DoD records matched on the respective combination; "0" indicates that there was no match; and a blank space indicates that no match was attempted.[1]

[1] The only exception to this convention is with respect to the Combination G match in the May 28 dataset, where a "0" can indicate both that no match was attempted and that no match was made. No match would have been attempted where the underlying voter registration record had matched on a Stage 1 or Stage 2 combination, or where the particular values in Combination G were not unique among all voter registration records to be compared.

15. No other matching comparisons between the Texas data and data from DMDC were undertaken beyond those set forth above, and in accordance with the steps provided in Ex. A.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on May 29, 2014.

Michelle Rudolph

Exhibit A

***United States v. Texas*: Federal Agency Algorithm Summary**

This document summarizes the database matches that both the United States and all other plaintiffs and the State of Texas have requested from Federal Agencies as part of the *Veasey v. Perry*/*United States v. Texas* litigation (S.D. Tex).

The matching process proceeds in three parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, the United States' one-to-many matches and the State of Texas's many-to-many matches are conducted between databases.

PART I: DATABASE PREPARATION

Stage 1: Extraction of Available Data from Federal Identification & Disability Databases

Step 1.1.1: Extract complete name into separate first name, middle name, and last name fields.

Step 1.1.2: Extract date of birth.

Step 1.1.3: Extract gender.

Step 1.1.4: Extract residential address and mailing address.

Step 1.1.5: Extract social security number.

Step 1.1.6: Extract Texas driver license number (only if present in Federal database).

Stage 2: Separate Valid Identification and Disability Records

Step 1.2.1: Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

Step 1.2.2: Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

Stage 3: Diagnostics

Step 1.3.1: Report the frequency of missing values for each field.

Step 1.3.2: Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.

Step 1.3.3: Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

Stage 4: Standardize Last Name

Step 1.4.1: Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.

Step 1.4.2: For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.

Step 1.4.3: Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.

Step 1.4.4: Code all missing values as blank fields.

Exhibit A

Stage 5: Standardize First Name and Middle Name

Step 1.5.1: Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.

Step 1.5.2: Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.[1]

Step 1.5.3: Code all missing values as blank fields.

Stage 6: Standardize Date of Birth

Step 1.6.1: Convert the date of birth to an eight-digit string of MMDDYYYY.

Step 1.6.2: Code all missing values as blank fields.

Stage 7: Standardize Gender

Step 1.7.1: Code gender as a string of 1 for females and 0 for males.

Step 1.7.2: Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.

Step 1.7.3: If missing values remain, code all missing values as blank fields.

[1] The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Stage 8: Standardize Address

Step 1.8.1: Convert the residential ZIP code to a string if it is stored as a numeric field.

Step 1.8.2: Where the residential address ZIP code is blank, populate that field with the value in the mailing address ZIP code field, if available.[2]

Step 1.8.3: Truncate the residential ZIP code field to the first five digits. *E.g.*, <77777-1234> becomes <77777>.

Step 1.8.4: Where the residential address field is blank, populate that field with the value in the mailing address field, if available.

Step 1.8.5: Where address field containing street address begins with a street number, isolate the street number. *E.g.*, <123 Main Street> becomes <123>.

Step 1.8.6: Where the address field begins with recognized strings indicating a mail box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444> becomes <444>.

Step 1.8.7: If missing values remain, code all missing values as blank fields.

Stage 9: Standardize Social Security Number

Step 1.9.1: Convert the social security number to a string if it is stored as a numeric field.

Step 1.9.2: Using full social security number, check for invalid SSNs. In the case of invalid SSNs, code as missing. *E.g.*, <123456789> becomes <>.

[2] For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Step 1.9.3: Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

Step 1.9.4: Code all missing values as blank fields.

PART II: DATABASE PREPARATION

Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps

Step 2.1.1: Create Combination A: First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.[3]

Step 2.1.2: Create Combination B: Last Name + Gender + DOB + Residential ZIP + Residential Street Number.

Step 2.1.3: Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.4: Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.5: Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.

Step 2.1.6: Create Combination F: First Name + Last Name + Gender + DOB.

Step 2.1.7: Create Combination M: Texas Driver License Number (where available).

[3] For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Exhibit A

Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps

Step 2.2.1: Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.[4]

Step 2.2.2: Create Combination H: SSN4 + Date of Birth + Residential ZIP.

Step 2.2.3: Create Combination I: SSN4 + First Name + Last Name + Date of Birth.

Step 2.2.4: Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.

Step 2.2.5: Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.

Step 2.2.6: Full Social Security Number.

Stage 3: Construct Identifiers Used Only For Texas's Many-to-Many Sweeps

Step 2.3.1: Create Combination for Texas's Sweep 1: SSN4 + Last Name + DOB.

Step 2.3.2: Create Combination for Texas's Sweep 3: First Name + Last Name + DOB

Step 2.3.3: Create Combination for Texas's Sweep 4: First Name + Middle Initial + Last Name + DOB

Note: Combinations for Texas's Sweeps 1 and 3 do not already exist as pre-made fields in the TEAM database extract but instead must be created from the underlying TEAM database fields, in addition to being constructed on the Federal database side. Texas's Sweep 4 is equivalent to

[4] Only for the State Department, create three further variations of Combination G created using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

the combination for the United States' Combination G. Texas's Sweep 2 is on full 9 social security number.[5]

Stage 4: Establish Identifier Uniqueness For Combinations A - L

Step 2.4.1: Generate a field that establishes the uniqueness of each identifier variable. For federal databases, for each combination A-L, generate a field that establishes uniqueness among only Texas records and a field that establishes uniqueness among nationwide records. *E.g.*, if only one record has the string <JEANSMITH01011950012377777> for Combination A, populate the uniqueness field for Combination A for that record as <1>. If four records have the string <JOHNSMITHA0101950> for Combination G, populate the uniqueness field for Combination G for each of those records as <2>, which indicates any number greater than one.

[5] For purposes of matching to U.S. Department of State Passport and Passport Card holder data, Texas's Sweep 3 and Sweep 4 are as follows: Sweep 3: Last Name + First Name (restricted to 32 characters) + DOB. Sweep 4: Last Name + First Name + Middle Initial (restricted to 32 characters) + DOB.

Exhibit A

PART III: MATCH DATABASES

Stage 1: United States' Primary One-to-Many Matching Sweeps

Step 3.1.1: For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identifier or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).

Step 3.1.2: Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identifier or disability database.

Step 3.1.3: Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identifier databases produced by the State of Texas.

Exhibit A

Stage 2: United States' Secondary One-to-Many Matching Sweeps

Step 3.2.1: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identifier or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.2.2: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number[6] in the TEAM database against the equivalent combination/field in the identifier or disability database.

Step 3.2.3: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identifier or disability database.

[6] The full social security number is not created as a separate "combination" as it is its own field stored within the TEAM database under the field name "ssn".

Step 3.2.4: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identifier or disability database.

Stage 3: United States' Nationwide Federal Sweeps

Step 3.3.1: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identifier or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.3.2: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identifier or disability database.

Step 3.3.3: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination

K against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.

Step 3.3.4: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination L against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.[7]

Stage 4: Texas' Many-to-Many Nationwide Sweeps

Step 3.4.1 Regardless of whether the combination for Sweep 1 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

[7] Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last_fix + First_fix + Middle_name from the TEAM database;
- Combination G2 to DOB + Last_fix + First_fix from the TEAM database;
- Combination G2 to DOB + Last_fix + First word of First_name from the TEAM database;
- Combination G3 to DOB + Last_fix + First_fix + Middle_Initial from the TEAM database;
- Combination G3 to DOB + Last_fix + First word of First_name + Middle_Initial from the TEAM database; and
- Combination G3 to DOB + Last_fix + First word of First_name + First character of Second word of First_fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

Step 3.4.2 Regardless of whether full 9 social security number is unique in the TEAM database, for Sweep 2, match against the equivalent field in a nationwide search of the Federal database.

Step 3.4.3 Regardless of whether the combination for Sweep 3 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Step 3.4.4 Regardless of whether the combination for Sweep 4 (Combination G) is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Note: For each of the Texas many-to-many sweeps:

- Indicate <1> if any TEAM combination matches a single combination in the Federal database
- Indicate <2> if any TEAM combination matches more than one record in the Federal database.
- Indicate <0> if no match is achieved.

Examples:

- If there are two TEAM records that have identical versions of the combination for Sweep 1, and there is one record in the Federal database that matches on this combination, both of the underlying TEAM records will have a matching output of <1> for Sweep 1.

- If there are three TEAM records that have identical versions of the combination for Sweep 3, and there are five records in the Federal database that match on that combination, the three TEAM records will each have a matching output of <2> for Sweep 3.

DMDC Step 1.3 Data Diagnostics

Data Element	Data Availability	"In Texas" Population	"Outside of Texas" Population
Last Name	available	746,010	8,285,517
	missing	0	1
First Name	available	746,004	8,285,426
	missing	6	92
Middle Name	available	669,476	7,656,604
	missing	76,534	628,914
SSN	available	746,010	8,285,517
	missing	0	1
Date of Birth	available	746,010	8,285,518
	missing	0	0
Gender	available	746,003	8,285,379
	missing	7	139
Address Text	available	746,010	8,108,935
	missing	0	176,583
Address State	available	746,010	8,026,896
	missing	0	258,622
Address Zip	available	745,993	8,072,943
	missing	17	212,575

Data Element	Data Quality*	"In Texas" Population	"Outside of Texas" Population
SSN	good	744,032	8,211,554
	bad	1,978	73,964
Date of Birth	good	745,936	8,284,603
	bad	74	915

NOTES:

An SSN was considered "bad" if it met any of the following criteria:

*Equaled 000000000, 111111111, 222222222, 333333333,…,999999999

*Equaled sequantial values 123456789 or 234567890

*The first 3 numbers were 000 or 666

*The first number was 9

*The middle node was 00 (ie xxx-00-xxxx)

*The last node was 0000 (ie xxx-xx-0000)

A Date of Birth was considered "bad" if it met either of the following criteria:

*Birth year was 1914 or prior (aka were 100 years of age or older)

*Representative of repeating digits (ie 01/01/01, 02/02/02,…,12/12/12)

Exhibit E

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*, Plaintiffs, v. RICK PERRY, *et al.*, Defendants.	Civil Action No. 2:13-cv-193 (NGR)
UNITED STATES OF AMERICA, Plaintiff, TEXAS LEAGUE OF YOUNG VOTERS EDUCATION FUND, *et al.*, Plaintiff-Intervenors, TEXAS ASSOCIATION OF HISPANIC COUNTY JUDGES AND COUNTY COMMISSIONERS, *et al.*, Plaintiff-Intervenors, v. STATE OF TEXAS, *et al.*, Defendants.	Civil Action No. 2:13-cv-263 (NGR)

TEXAS STATE CONFERENCE OF NAACP BRANCHES, *et al.*, Plaintiffs, v. JOHN STEEN, *et al.*, Defendants.	Civil Action No. 2:13-cv-291 (NGR)
BELINDA ORTIZ, *et al.*, Plaintiffs, v. STATE OF TEXAS, *et al.*, Defendants	Civil Action No. 2:13-cv-348 (NGR)

DECLARATION OF YEMI B. OSHINNAIYE

I, Yemi B. Oshinnaiye, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as Acting Associate Chief of Systems Operations, Office of Information Technology ("OIT"), U.S. Citizenship and Immigration Services ("USCIS"), Department of Homeland Security. I have served in this capacity since May 15, 2014. My duties include managing Branch and Project Managers that oversee development, deployment and production support for systems that support USCIS current operations. Prior to this position I served as Branch Chief for Benefits and Biometrics, USCIS OIT. I served in that position beginning in October, 2012. My duties included Management of Project/Program managers that oversaw the development, deployment and production support for systems that supported the administration, management and granting of

benefits from USCIS and systems that validate eligibility for those benefits. I have personal knowledge of the information contained in this declaration based upon my work for the USCIS, and my personal involvement in and supervision of the database comparison described below.

2. On February 27, 2014, my agency received an encrypted hard drive containing data regarding Texas registered voters from the United States Department of Justice ("DOJ") in the form of a 7 gigabyte .csv file containing 13,564,420 records for Texas registered voters. USCIS was able to load all 13,564,420 records.

3. The Texas data that USCIS received contained the following fields:

- Vuid
- last_name
- first_name
- middle_name
- ssn
- ssn4_str
- gender_code
- official_id
- dl_id_str
- date_of_birth
- voter_status_code
- block_number
- street_name
- city
- state
- zip_code
- zip_str
- mail_adrs_1
- last_fix
- multi_lastname1
- multi_lastname2
- first_fix
- middle_initial
- fem_str
- dob_str
- zip5_res_alt
- str_num_res_alt
- AGDN
- AGDN_TAG
- AGDN_unique
- AGDNlast
- AGDNlast_TAG
- AGDNlast_unique
- AGD
- AGD_TAG
- AGD_unique
- ADN
- ADN_TAG
- ADN_unique
- AGN
- AGN_TAG
- AGN_unique
- GDN
- GDN_TAG
- GDN_unique
- DLN
- DLN_TAG
- DLN_unique
- unique_count
- FMLD
- FMLD_TAG
- FMLD_unique

- SDZ
- SDZ_TAG
- SDZ_unique
- SND
- SND_TAG
- SND_unique
- FML1D
- FML1D_TAG
- FML1D_unique
- FML2D
- FML2D_TAG
- FML2D_unique
- unique_count2
- var_fill
- K2
- K3
- L2
- L3

4. USCIS executed the data preparation and comparison algorithm as it was relevant to USCIS databases containing U.S. Certificate of Citizenship or U.S. Certificate of Naturalization data , attached as Ex. A, that was provided by DOJ.[1]

5. Through this process, USCIS attempted to match particular combinations of identifying information for Texas registered voters (*e.g.*, first and last name, gender, and date of birth) with the same or related combinations of identifying information with respect to holders of a U.S. Certificate of Citizenship or a U.S. Certificate of Naturalization found in the USCIS Computer Linked Application Information Management System 4 ("CLAIMS 4") and Central Index System ("CIS"), which contain U.S. Certificate of Citizenship or U.S. Certificate of Naturalization data. For each particular combination, USCIS identified instances where the identifying information from the Texas registered voter information matched with the same or related combinations of identifying information for one or more records of individuals issued a U.S. Certificate of Citizenship or a U.S. Certificate of Naturalization. The frequencies of any missing values in the underlying USCIS data are provided in Ex. B.

6. For each sweep through the relevant USCIS dataset, we tracked the results,

[1] The only data preparation step listed in Ex. A that USCIS was unable to complete was the Step 1.7.2 imputation of gender for USCIS records with addresses outside of Texas. No gender imputation was conducted for USCIS records with an address outside of Texas that contained a blank gender value.

on a record-by-record basis, of the database comparisons on all of the requested combinations. In order to speed computer processing time, the results were appended only to a unique identifier, the Var_Fill value, that had been provided for each Texas voter registration record we received. No other underlying data from the voter registration database was included in the results file sent to DOJ.

7. On May 23, 2014, USCIS transferred responsive data to DOJ, in the form of a 535 megabyte text (.txt) file, with comma delimiters.

8. The results of each sweep, as described in Ex. A, can be found in the following columns of the data USCIS returned to DOJ:

   a. "USA" contains results of the Stage 1, Step 3.1.1, Combination A match;

   b. "USB" contains results of the Stage 1, Step 3.1.2, Combination B match;

   c. "USC" contains results of the Stage 1, Step 3.1.2, Combination C match;

   d. "USD" contains results of the Stage 1, Step 3.1.2, Combination D match;

   e. "USE" contains results of the Stage 1, Step 3.1.2, Combination E match;

   f. "USF" contains results of the Stage 1, Step 3.1.2, Combination F match;

   g. "USG" contains results of the Stage 2, Step 3.2.1, Combination G match;

   h. "USH" contains results of the Stage 2, Step 3.2.2, Combination H match;

   i. "USI" contains results of the Stage 2, Step 3.2.2, Combination I match;

   j. "USSSN" contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;

   k. "USGK" contains results of the Stage 2, Step 3.2.3, Combination K to G match;

   l. "USK" contains results of the Stage 2, Step 3.2.3, Combination K match;

   m. "USKL" contains results of the Stage 2, Step 3.2.3, Combination K to L match;

n. "USGL" contains results of the Stage 2, Step 3.2.4, Combination L to G match;

o. "USL" contains results of the Stage 2, Step 3.2.4, Combination L match;

p. "USLK" contains results of the Stage 2, Step 3.2.4, Combination L to K match;

q. "USF-NW" contains results of the Stage 3, Step 3.3.1, Combination F match;

r. "USG-NW" contains results of the Stage 3, Step 3.3.2, Combination G match;

s. "USI-NW" contains results of the Stage 3, Step 3.3.2, Combination I match;

t. "USSSN-NW" contains results of the Stage 3, Step 3.3.2, nine-digit social security number match;

u. "USGK_NW" contains results of the Stage 3, Step 3.3.3, Combination K to G match;

v. "USK_NW" contains results of the Stage 2, Step 3.3.3, Combination K match;

w. "USKL_NW" contains results of the Stage 2, Step 3.3.3, Combination K to L match;

x. "USGL_NW" contains results of the Stage 3, Step 3.3.4, Combination L to G match;

y. "USL_NW" contains results of the Stage 3, Step 3.3.4, Combination L match;

z. "USLK_NW" contains results of the Stage 3, Step 3.3.4, Combination L to K match;

aa. "Texas1" contains results of the Stage 4, Step 3.4.1, Sweep 1 match;

bb. "Texas2" contains results of the Stage 4, Step 3.4.2, Sweep 2 match;

cc. "Texas3" contains results of the Stage 4, Step 3.4.3, Sweep 3 match; and

dd. "Texas4" contains results of the Stage 4, Step 3.4.4, Sweep 4 match.

9. No other matching comparisons between the Texas data and data from USCIS were undertaken beyond those set forth in Ex. A.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on May 29, 2014.

YEMI B. OSHINNAIYE

Acting Associate Chief of Systems Operations, Office of Information Technology, U.S. Citizenship and Immigration Services

Exhibit A

***United States v. Texas*: Federal Agency Algorithm Summary**

This document summarizes the database matches that both the United States and all other plaintiffs and the State of Texas have requested from Federal Agencies as part of the *Veasey v. Perry*/*United States v. Texas* litigation (S.D. Tex).

The matching process proceeds in three parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, the United States' one-to-many matches and the State of Texas's many-to-many matches are conducted between databases.

PART I: DATABASE PREPARATION

Stage 1: Extraction of Available Data from Federal Identification & Disability Databases

Step 1.1.1: Extract complete name into separate first name, middle name, and last name fields.

Step 1.1.2: Extract date of birth.

Step 1.1.3: Extract gender.

Step 1.1.4: Extract residential address and mailing address.

Step 1.1.5: Extract social security number.

Step 1.1.6: Extract Texas driver license number (only if present in Federal database).

Stage 2: Separate Valid Identification and Disability Records

Step 1.2.1: Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

Step 1.2.2: Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

Stage 3: Diagnostics

Step 1.3.1: Report the frequency of missing values for each field.

Step 1.3.2: Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.

Step 1.3.3: Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

Stage 4: Standardize Last Name

Step 1.4.1: Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.

Step 1.4.2: For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.

Step 1.4.3: Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.

Step 1.4.4: Code all missing values as blank fields.

Exhibit A

Stage 5: Standardize First Name and Middle Name

Step 1.5.1: Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.

Step 1.5.2: Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.[1]

Step 1.5.3: Code all missing values as blank fields.

Stage 6: Standardize Date of Birth

Step 1.6.1: Convert the date of birth to an eight-digit string of MMDDYYYY.

Step 1.6.2: Code all missing values as blank fields.

Stage 7: Standardize Gender

Step 1.7.1: Code gender as a string of 1 for females and 0 for males.

Step 1.7.2: Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.

Step 1.7.3: If missing values remain, code all missing values as blank fields.

[1] The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Stage 8: Standardize Address

Step 1.8.1: Convert the residential ZIP code to a string if it is stored as a numeric field.

Step 1.8.2: Where the residential address ZIP code is blank, populate that field with the value in the mailing address ZIP code field, if available.[2]

Step 1.8.3: Truncate the residential ZIP code field to the first five digits. *E.g.*, <77777-1234> becomes <77777>.

Step 1.8.4: Where the residential address field is blank, populate that field with the value in the mailing address field, if available.

Step 1.8.5: Where address field containing street address begins with a street number, isolate the street number. *E.g.*, <123 Main Street> becomes <123>.

Step 1.8.6: Where the address field begins with recognized strings indicating a mail box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444> becomes <444>.

Step 1.8.7: If missing values remain, code all missing values as blank fields.

Stage 9: Standardize Social Security Number

Step 1.9.1: Convert the social security number to a string if it is stored as a numeric field.

Step 1.9.2: Using full social security number, check for invalid SSNs. In the case of invalid SSNs, code as missing. *E.g.*, <123456789> becomes <>.

[2] For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Step 1.9.3: Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

Step 1.9.4: Code all missing values as blank fields.

PART II: DATABASE PREPARATION

Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps

Step 2.1.1: Create Combination A: First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.[3]

Step 2.1.2: Create Combination B: Last Name + Gender + DOB + Residential ZIP + Residential Street Number.

Step 2.1.3: Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.4: Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.5: Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.

Step 2.1.6: Create Combination F: First Name + Last Name + Gender + DOB.

Step 2.1.7: Create Combination M: Texas Driver License Number (where available).

[3] For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Exhibit A

Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps

Step 2.2.1: Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.[4]

Step 2.2.2: Create Combination H: SSN4 + Date of Birth + Residential ZIP.

Step 2.2.3: Create Combination I: SSN4 + First Name + Last Name + Date of Birth.

Step 2.2.4: Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.

Step 2.2.5: Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.

Step 2.2.6: Full Social Security Number.

Stage 3: Construct Identifiers Used Only For Texas's Many-to-Many Sweeps

Step 2.3.1: Create Combination for Texas's Sweep 1: SSN4 + Last Name + DOB.

Step 2.3.2: Create Combination for Texas's Sweep 3: First Name + Last Name + DOB

Step 2.3.3: Create Combination for Texas's Sweep 4: First Name + Middle Initial + Last Name + DOB

Note: Combinations for Texas's Sweeps 1 and 3 do not already exist as pre-made fields in the TEAM database extract but instead must be created from the underlying TEAM database fields, in addition to being constructed on the Federal database side. Texas's Sweep 4 is equivalent to

[4] Only for the State Department, create three further variations of Combination G created using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

the combination for the United States' Combination G. Texas's Sweep 2 is on full 9 social security number.[5]

Stage 4: Establish Identifier Uniqueness For Combinations A - L

Step 2.4.1: Generate a field that establishes the uniqueness of each identifier variable. For federal databases, for each combination A-L, generate a field that establishes uniqueness among only Texas records and a field that establishes uniqueness among nationwide records. *E.g.*, if only one record has the string <JEANSMITH01011950012377777> for Combination A, populate the uniqueness field for Combination A for that record as <1>. If four records have the string <JOHNSMITHA0101950> for Combination G, populate the uniqueness field for Combination G for each of those records as <2>, which indicates any number greater than one.

[5] For purposes of matching to U.S. Department of State Passport and Passport Card holder data, Texas's Sweep 3 and Sweep 4 are as follows: Sweep 3: Last Name + First Name (restricted to 32 characters) + DOB. Sweep 4: Last Name + First Name + Middle Initial (restricted to 32 characters) + DOB.

Exhibit A

PART III: MATCH DATABASES

Stage 1: United States' Primary One-to-Many Matching Sweeps

Step 3.1.1: For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identifier or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).

Step 3.1.2: Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identifier or disability database.

Step 3.1.3: Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identifier databases produced by the State of Texas.

Exhibit A

Stage 2: United States' Secondary One-to-Many Matching Sweeps

Step 3.2.1: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identifier or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.2.2: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number[6] in the TEAM database against the equivalent combination/field in the identifier or disability database.

Step 3.2.3: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identifier or disability database.

[6] The full social security number is not created as a separate "combination" as it is its own field stored within the TEAM database under the field name "ssn".

Exhibit A

Step 3.2.4: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identifier or disability database.

Stage 3: United States' Nationwide Federal Sweeps

Step 3.3.1: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identifier or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.3.2: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identifier or disability database.

Step 3.3.3: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination

K against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.

Step 3.3.4: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination L against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.[7]

Stage 4: Texas' Many-to-Many Nationwide Sweeps

Step 3.4.1 Regardless of whether the combination for Sweep 1 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

[7] Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last_fix + First_fix + Middle_name from the TEAM database;
- Combination G2 to DOB + Last_fix + First_fix from the TEAM database;
- Combination G2 to DOB + Last_fix + First word of First_name from the TEAM database;
- Combination G3 to DOB + Last_fix + First_fix + Middle_Initial from the TEAM database;
- Combination G3 to DOB + Last_fix + First word of First_name + Middle_Initial from the TEAM database; and
- Combination G3 to DOB + Last_fix + First word of First_name + First character of Second word of First_fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

Step 3.4.2 Regardless of whether full 9 social security number is unique in the TEAM database, for Sweep 2, match against the equivalent field in a nationwide search of the Federal database.

Step 3.4.3 Regardless of whether the combination for Sweep 3 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Step 3.4.4 Regardless of whether the combination for Sweep 4 (Combination G) is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Note: For each of the Texas many-to-many sweeps:

- Indicate <1> if any TEAM combination matches a single combination in the Federal database
- Indicate <2> if any TEAM combination matches more than one record in the Federal database.
- Indicate <0> if no match is achieved.

Examples:

- If there are two TEAM records that have identical versions of the combination for Sweep 1, and there is one record in the Federal database that matches on this combination, both of the underlying TEAM records will have a matching output of <1> for Sweep 1.

- If there are three TEAM records that have identical versions of the combination for Sweep 3, and there are five records in the Federal database that match on that combination, the three TEAM records will each have a matching output of <2> for Sweep 3.

Exhibit B

United States Citizenship and Immigration Services

Step 1.3 Diagnostics

USCIS had a population of 19,401,814 United States Citizens who were presented either the N550 Naturalization Certificate or the N560 Certificate of Citizenship. This listing was extracted from the USCIS Central Index System which is the most comprehensive listing of United States Citizens.

Below are the numbers of USCIS records analyzed missing information from a particular listed field.

Step 1.3.1

Missing first name – 49,814
Missing middle name – 7,180,037
Missing last name – 0
Missing date of birth – 18,632
Missing gender – 8,251,075
Missing SSN – 14,264,377

Step 1.3.2:

Number of Records with Invalid SSNs:

5	123456789
2	987654321
12	111111111
3	555555555
2	666666666
1	777777777
2	888888888
40	999999999

Step 1.3.3

Number of Records with Potentially Invalid dates of birth

85	19010101
49	19020202
29	19030303
41	19040404
59	19050505
44	19060606
55	19070707
58	19080808
69	19090909
103	19111111
122	19121212

Exhibit F

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*, Plaintiffs, v. RICK PERRY, *et al.*, Defendants.	Civil Action No. 2:13-cv-193 (NGR)
UNITED STATES OF AMERICA, Plaintiff, TEXAS LEAGUE OF YOUNG VOTERS EDUCATION FUND, *et al.*, Plaintiff-Intervenors, TEXAS ASSOCIATION OF HISPANIC COUNTY JUDGES AND COUNTY COMMISSIONERS, *et al.*, Plaintiff-Intervenors, v. STATE OF TEXAS, *et al.*, Defendants.	Civil Action No. 2:13-cv-263 (NGR)

TEXAS STATE CONFERENCE OF NAACP BRANCHES, *et al.*, Plaintiffs, v. JOHN STEEN, *et al.*, Defendants.	Civil Action No. 2:13-cv-291 (NGR)
BELINDA ORTIZ, *et al.*, Plaintiffs, v. STATE OF TEXAS, *et al.*, Defendants	Civil Action No. 2:13-cv-348 (NGR)

DECLARATION OF STEVE STRAUSLER

I, Steve Strausler, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as Computer Specialist with the Social Security Administration ("SSA"). I have served in this capacity for 33 years. My duties include updating and maintenance of SSA's Master Beneficiary Record, System of Records Number 60-0090 (published at 71 F.R. 1826 (January 11, 2006)).
2. I have personal knowledge of the information contained in this declaration based upon my work for SSA, and my personal involvement in completion of the database comparison described below.

3. On February 28, 2014, SSA received from the United States Department of Justice ("DOJ") an encrypted hard drive containing data regarding Texas registered voters in the form of a 7 gigabyte .csv file containing 13,564,420 records for Texas registered voters. SSA was unable to use this variable width file. We requested that the DOJ provide a fixed width file, which we received from DOJ on April 3, 2014. 13,564,419 records were provided in the fixed width file and SSA was able to load all of them.[1] In order to hold down the file size and make the fixed width file more manageable for SSA to load, this file contained only the fields that were strictly necessary to conduct the database comparison process, rather than all fields that were used to construct the matching combinations, as had been sent in the initial variable width file.
4. The Texas data on the 9.4 GB fixed width file that SSA received contained the following fields:

[1] One record from the variable width file was omitted because of data formatting issues with that record.

- vuid
- ssn
- ssn4_str
- zip_str
- last_fix
- first_fix
- middle_initial
- dob_str
- AGDN
- AGDNlast
- AGD
- ADN
- AGN
- GDN
- DLN
- FMLD
- SDZ
- SND
- FMLID
- FML2D
- var_fill
- K2
- K3
- L2
- L3

5. SSA executed the data preparation and comparison steps, attached as Ex. A, that were provided by the DOJ.

6. Through this process, SSA attempted to match particular combinations of identifying information for Texas registered voters (*e.g.*, first and last name, gender, and date of birth) with the same or related combinations of identifying information with respect to records from SSA's master beneficiary records[2] of persons who are 18 years of age or older, and who SSA has determined to be in current payment status for disability.[3]

7. For each sweep through the relevant SSA data, we appended results to the Texas data to indicate, on a record-by-record basis, the output of the database comparisons on all of the requested combinations. The frequencies of missing values in the underlying SSA data are provided as Ex. B.

---

[2] SSA's Master Beneficiary Records includes records from the following SSA systems of records: (1) Master Beneficiary Record, System of Record Notice 60-0090 (last published at 71 F.R. 1826 (January 11, 2006)) and (2) Supplemental Security Income Record and Special Veterans Benefits, System of Record Notice 60-0103 (last published at 71 F.R. 1830 (January 11, 2006)).

[3] For Supplemental Security Income (SSI) purposes, SSA's search included individuals who are in active disability status, in addition to individuals who have been determined to be blind.

8. On May 28, 2014, SSA transferred responsive data to the Department of Justice, in the form of a 9.75 gigabyte, fixed width, encrypted text file. The name of the results file transferred to the Department of Justice is texas.dojfinal.0527.[4]

9. The results of each sweep, as described in Ex. A, can be found in the following fixed positions in the data that SSA returned to the Department of Justice on May 28:

   a. Position 741 contains results of the Stage 1, Step 3.1.1, Combination A match;

   b. Position 742 contains results of the Stage 1, Step 3.1.2, Combination B match;

   c. Position 743 contains results of the Stage 1, Step 3.1.2, Combination C match;

   d. Position 744 contains results of the Stage 1, Step 3.1.2, Combination D match;

   e. Position 745 contains results of the Stage 1, Step 3.1.2, Combination E match;

   f. Position 746 contains results of the Stage 1, Step 3.1.2, Combination F match;

   g. Position 747 contains results of the Stage 2, Step 3.2.1, Combination G match;

   h. Position 748 contains results of the Stage 2, Step 3.2.2, Combination H match;

   i. Position 749 contains results of the Stage 2, Step 3.2.2, Combination I match;

   j. Position 756 contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;

   k. Position 750 contains results of the Stage 2, Step 3.2.3, Combination K to G match;

   l. Position 751 contains results of the Stage 2, Step 3.2.3, Combination K match;

   m. Position 752 contains results of the Stage 2, Step 3.2.3, Combination K to L match;

   n. Position 753 contains results of the Stage 2, Step 3.2.4, Combination L to G match;

   o. Position 755 contains results of the Stage 2, Step 3.2.4, Combination L match;

[4] SSA transferred identical results files to the Department of Justice on May 16, 2014 and May 23, 2014. These files could not be accessed by Department of Justice staff because of problems unencrypting the file.

p. Position 754 contains results of the Stage 2, Step 3.2.4, Combination L to K match;

q. Position 757 contains results of the Stage 3, Step 3.3.1, Combination F match;

r. Position 758 contains results of the Stage 3, Step 3.3.2, Combination G match;

s. Position 759 contains results of the Stage 3, Step 3.3.2, Combination I match;

t. Position 766 contains results of the Stage 3, Step 3.3.2, nine-digit social security number match;

u. Position 760 contains results of the Stage 3, Step 3.3.3, Combination K to G match;

v. Position 761 contains results of the Stage 2, Step 3.3.3, Combination K match;

w. Position 762 contains results of the Stage 2, Step 3.3.3, Combination K to L match;

x. Position 763 contains results of the Stage 3, Step 3.3.4, Combination L to G match;

y. Position 765 contains results of the Stage 3, Step 3.3.4, Combination L match;

z. Position 764 contains results of the Stage 3, Step 3.3.4, Combination L to K match;

aa. Position 767 contains results of the Stage 4, Step 3.4.1, Sweep 1 match;

bb. Position 768 contains results of the Stage 4, Step 3.4.2, Sweep 2 match;

cc. Position 769 contains results of the Stage 4, Step 3.4.3, Sweep 3 match; and

dd. Position 770 contains results of the Stage 4, Step 3.4.4, Sweep 4 match.

10. For each of these positions, "1" indicates that the respective combination from one SSA record matched; "2" indicates that two or more SSA records matched on the respective combination; "0" indicates that there was no match; and a blank space indicates that no match was attempted.

11. When SSA initially ran the data comparison process, the results in Positions 750, 751, and 752 were grouped together rather than reported separately, as were the results in Positions 753, 754, and 755; Positions 760, 761, and 762; and Positions 763, 764, and

765. SSA did not produce those combined results to DOJ. Instead, at DOJ's request, SSA separated those results into separate positions, as reflected in the May 28 file. Likewise, when SSA initially ran the data comparison results for the Texas Stage 4 matches, only those SSA records that had not been matched on any of the Stage 1 to 3 matches were included in the eligible-for-matching population. SSA did not produce those results to DOJ. Instead, at DOJ's request, SSA re-ran the Texas Stage 4 matches, comparing the voter registration data to all eligible SSA records nationwide (as described in paragraph 6 above), regardless of whether they had been matched on any of the Stage 1 to 3 sweeps.

12. No other matching comparisons between the Texas data and data from the SSA were undertaken beyond those set forth above, and in accordance with the steps provided in Ex. A.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on May 29, 2014.

Steve Strausler

Exhibit A

***United States v. Texas*: Federal Agency Algorithm Summary**

This document summarizes the database matches that both the United States and all other plaintiffs and the State of Texas have requested from Federal Agencies as part of the *Veasey v. Perry*/*United States v. Texas* litigation (S.D. Tex).

The matching process proceeds in three parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, the United States' one-to-many matches and the State of Texas's many-to-many matches are conducted between databases.

PART I: DATABASE PREPARATION

Stage 1: Extraction of Available Data from Federal Identification & Disability Databases

Step 1.1.1: Extract complete name into separate first name, middle name, and last name fields.

Step 1.1.2: Extract date of birth.

Step 1.1.3: Extract gender.

Step 1.1.4: Extract residential address and mailing address.

Step 1.1.5: Extract social security number.

Step 1.1.6: Extract Texas driver license number (only if present in Federal database).

Stage 2: Separate Valid Identification and Disability Records

Step 1.2.1: Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

Exhibit A

Step 1.2.2: Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

Stage 3: Diagnostics

Step 1.3.1: Report the frequency of missing values for each field.

Step 1.3.2: Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.

Step 1.3.3: Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

Stage 4: Standardize Last Name

Step 1.4.1: Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.

Step 1.4.2: For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.

Step 1.4.3: Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.

Step 1.4.4: Code all missing values as blank fields.

Exhibit A

Stage 5: Standardize First Name and Middle Name

Step 1.5.1: Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.

Step 1.5.2: Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.[1]

Step 1.5.3: Code all missing values as blank fields.

Stage 6: Standardize Date of Birth

Step 1.6.1: Convert the date of birth to an eight-digit string of MMDDYYYY.

Step 1.6.2: Code all missing values as blank fields.

Stage 7: Standardize Gender

Step 1.7.1: Code gender as a string of 1 for females and 0 for males.

Step 1.7.2: Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.

Step 1.7.3: If missing values remain, code all missing values as blank fields.

[1] The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Stage 8: Standardize Address

Step 1.8.1: Convert the residential ZIP code to a string if it is stored as a numeric field.

Step 1.8.2: Where the residential address ZIP code is blank, populate that field with the value in the mailing address ZIP code field, if available.[2]

Step 1.8.3: Truncate the residential ZIP code field to the first five digits. *E.g.*, <77777-1234> becomes <77777>.

Step 1.8.4: Where the residential address field is blank, populate that field with the value in the mailing address field, if available.

Step 1.8.5: Where address field containing street address begins with a street number, isolate the street number. *E.g.*, <123 Main Street> becomes <123>.

Step 1.8.6: Where the address field begins with recognized strings indicating a mail box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444> becomes <444>.

Step 1.8.7: If missing values remain, code all missing values as blank fields.

Stage 9: Standardize Social Security Number

Step 1.9.1: Convert the social security number to a string if it is stored as a numeric field.

Step 1.9.2: Using full social security number, check for invalid SSNs. In the case of invalid SSNs, code as missing. *E.g.*, <123456789> becomes <>.

[2] For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Step 1.9.3: Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

Step 1.9.4: Code all missing values as blank fields.

PART II: DATABASE PREPARATION

Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps

Step 2.1.1: Create Combination A: First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.[3]

Step 2.1.2: Create Combination B: Last Name + Gender + DOB + Residential ZIP + Residential Street Number.

Step 2.1.3: Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.4: Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.5: Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.

Step 2.1.6: Create Combination F: First Name + Last Name + Gender + DOB.

Step 2.1.7: Create Combination M: Texas Driver License Number (where available).

[3] For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Exhibit A

Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps

Step 2.2.1: Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.[4]

Step 2.2.2: Create Combination H: SSN4 + Date of Birth + Residential ZIP.

Step 2.2.3: Create Combination I: SSN4 + First Name + Last Name + Date of Birth.

Step 2.2.4: Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.

Step 2.2.5: Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.

Step 2.2.6: Full Social Security Number.

Stage 3: Construct Identifiers Used Only For Texas's Many-to-Many Sweeps

Step 2.3.1: Create Combination for Texas's Sweep 1: SSN4 + Last Name + DOB.

Step 2.3.2: Create Combination for Texas's Sweep 3: First Name + Last Name + DOB

Step 2.3.3: Create Combination for Texas's Sweep 4: First Name + Middle Initial + Last Name + DOB

Note: Combinations for Texas's Sweeps 1 and 3 do not already exist as pre-made fields in the TEAM database extract but instead must be created from the underlying TEAM database fields, in addition to being constructed on the Federal database side. Texas's Sweep 4 is equivalent to

[4] Only for the State Department, create three further variations of Combination G created using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

the combination for the United States' Combination G. Texas's Sweep 2 is on full 9 social security number.[5]

Stage 4: Establish Identifier Uniqueness For Combinations A - L

Step 2.4.1: Generate a field that establishes the uniqueness of each identifier variable. For federal databases, for each combination A-L, generate a field that establishes uniqueness among only Texas records and a field that establishes uniqueness among nationwide records. *E.g.*, if only one record has the string <JEANSMITH01011950012377777> for Combination A, populate the uniqueness field for Combination A for that record as <1>. If four records have the string <JOHNSMITHA0101950> for Combination G, populate the uniqueness field for Combination G for each of those records as <2>, which indicates any number greater than one.

[5] For purposes of matching to U.S. Department of State Passport and Passport Card holder data, Texas's Sweep 3 and Sweep 4 are as follows: Sweep 3: Last Name + First Name (restricted to 32 characters) + DOB. Sweep 4: Last Name + First Name + Middle Initial (restricted to 32 characters) + DOB.

Exhibit A

PART III: MATCH DATABASES

Stage 1: United States' Primary One-to-Many Matching Sweeps

Step 3.1.1: For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identifier or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).

Step 3.1.2: Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identifier or disability database.

Step 3.1.3: Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identifier databases produced by the State of Texas.

Exhibit A

Stage 2: United States' Secondary One-to-Many Matching Sweeps

Step 3.2.1: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identifier or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.2.2: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number[6] in the TEAM database against the equivalent combination/field in the identifier or disability database.

Step 3.2.3: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identifier or disability database.

[6] The full social security number is not created as a separate "combination" as it is its own field stored within the TEAM database under the field name "ssn".

Exhibit A

Step 3.2.4: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identifier or disability database.

Stage 3: United States' Nationwide Federal Sweeps

Step 3.3.1: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identifier or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.3.2: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identifier or disability database.

Step 3.3.3: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination

K against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.

Step 3.3.4: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination L against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.[7]

Stage 4: Texas' Many-to-Many Nationwide Sweeps

Step 3.4.1 Regardless of whether the combination for Sweep 1 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

[7] Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last_fix + First_fix + Middle_name from the TEAM database;
- Combination G2 to DOB + Last_fix + First_fix from the TEAM database;
- Combination G2 to DOB + Last_fix + First word of First_name from the TEAM database;
- Combination G3 to DOB + Last_fix + First_fix + Middle_Initial from the TEAM database;
- Combination G3 to DOB + Last_fix + First word of First_name + Middle_Initial from the TEAM database; and
- Combination G3 to DOB + Last_fix + First word of First_name + First character of Second word of First_fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

Step 3.4.2 Regardless of whether full 9 social security number is unique in the TEAM database, for Sweep 2, match against the equivalent field in a nationwide search of the Federal database.

Step 3.4.3 Regardless of whether the combination for Sweep 3 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Step 3.4.4 Regardless of whether the combination for Sweep 4 (Combination G) is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Note: For each of the Texas many-to-many sweeps:

- Indicate <1> if any TEAM combination matches a single combination in the Federal database
- Indicate <2> if any TEAM combination matches more than one record in the Federal database.
- Indicate <0> if no match is achieved.

Examples:

- If there are two TEAM records that have identical versions of the combination for Sweep 1, and there is one record in the Federal database that matches on this combination, both of the underlying TEAM records will have a matching output of <1> for Sweep 1.

- If there are three TEAM records that have identical versions of the combination for Sweep 3, and there are five records in the Federal database that match on that combination, the three TEAM records will each have a matching output of <2> for Sweep 3.

Exhibit B

Social Security Administration

Step 1.3 Diagnostics on SSA Data

Step 1.3.1:

Missing first name – 0%
Missing middle name – 20.944%
Missing last name – 0%
Missing date of birth – 0%
Missing gender – 0%
Missing residential address – 63.938%
Missing mailing address – 0%
Missing SSN – 0.0264%

Step 1.3.2:

Invalid SSN – 0%

Step 1.3.3

Invalid date of birth – 0%

Exhibit G

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*, Plaintiffs, v. RICK PERRY, *et al.*, Defendants.	Civil Action No. 2:13-cv-193 (NGR)
UNITED STATES OF AMERICA, Plaintiff, TEXAS LEAGUE OF YOUNG VOTERS EDUCATION FUND, *et al.*, Plaintiff-Intervenors, TEXAS ASSOCIATION OF HISPANIC COUNTY JUDGES AND COUNTY COMMISSIONERS, *et al.*, Plaintiff-Intervenors, v. STATE OF TEXAS, *et al.*, Defendants.	Civil Action No. 2:13-cv-263 (NGR)

TEXAS STATE CONFERENCE OF NAACP BRANCHES, *et al.*, Plaintiffs, v. JOHN STEEN, *et al.*, Defendants.	Civil Action No. 2:13-cv-291 (NGR)
BELINDA ORTIZ, *et al.*, Plaintiffs, v. STATE OF TEXAS, *et al.*, Defendants	Civil Action No. 2:13-cv-348 (NGR)

DECLARATION OF KENNETH SMITH

I, Kenneth Smith, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as the Assistant Director in the Veterans Benefits Administration's (VBA) Office of Performance Analysis and Integrity. I have served in this capacity for three years, nine months. My duties include the collection and reporting of VBA business and operational data sourced from the Enterprise Data Warehouse.
2. I have personal knowledge of the information contained in this declaration based upon my work for the Veterans Benefits Administration ("VBA") at the U.S. Department of Veterans Affairs ("VA"), and my supervision of the completion of the database comparison described below.

3. On March 5, 2014, the VA received from the United States Department of Justice an encrypted hard drive containing data regarding Texas registered voters in the form of a 7 gigabyte .csv file containing 13,564,420 records for Texas registered voters. VBA was able to load 13,555,657 of the 13,564,420 records. The 8,763 voter records that VBA was not able to load and analyze as part of the database comparison process (0.0006% of the Texas voter registration records received) contained embedded commas in an address or name field. These embedded commas prevented VBA from being able to load and analyze only those affected records.

4. The Texas data that VBA received contained the following fields:

- Vuid
- last_name
- first_name
- middle_name
- ssn
- ssn4_str
- gender_code
- official_id
- dl_id_str
- date_of_birth
- voter_status_code
- block_number
- street_name
- city
- state
- zip_code
- zip_str
- mail_adrs_1
- last_fix
- multi_lastname1
- multi_lastname2
- first_fix
- middle_initial
- fem_str
- dob_str
- zip5_res_alt
- str_num_res_alt
- AGDN
- AGDN_TAG
- AGDN_unique
- AGDNlast
- AGDNlast_TAG
- AGDNlast_unique
- AGD
- AGD_TAG
- AGD_unique
- and
- ADN_TAG
- ADN_unique
- AGN
- AGN_TAG
- AGN_unique
- GDN
- GDN_TAG
- GDN_unique
- DLN
- DLN_TAG
- DLN_unique
- unique_count
- FMLD
- FMLD_TAG
- FMLD_unique

- SDZ
- SDZ_TAG
- SDZ_unique
- SND
- SND_TAG
- SND_unique
- FML1D
- FML1D_TAG
- FML1D_unique
- FML2D
- FML2D_TAG
- FML2D_unique
- unique_count2
- var_fill
- K2
- K3
- L2
- L3

5. VBA executed the data preparation and comparison steps, attached as Ex. A, that were provided by the United States Department of Justice.

6. Through this process, VBA attempted to match particular combinations of identifying information for Texas registered voters (*e.g.*, first and last name, gender, and date of birth) with the same or related combinations of identifying information with respect to records for living Veterans with a 50% or higher disability rating, either service connected or non-service connected. For each particular combination, VBA identified instances where the identifying information for a Texas registered voter matched with the respective combination of identifying information for one or more living Veterans with a disability rating of 50% or higher.

7. For each sweep through the relevant VBA datasets, we appended a column to the Texas data to indicate, on a record-by-record basis, the output of the database comparisons on all of the requested combinations. The frequencies of any missing values in the underlying VBA data are provided as Ex. B.

8. On May 5, 2014, VBA transferred responsive data to the Department of Justice, in the form of a 3.2 gigabyte, zipped variable width file with pipe delimiters. The unzipped file is 37.8 gigabytes. The name of the results file transferred to the Department of Justice is tx_folder_project.texas_results_table_final_output.

9. The results of each sweep, as described in Ex. A, can be found in the following columns of the data that VBA returned to the Department of Justice on May 5:

   a. "USA_MATCHES_ID" contains results of the Stage 1, Step 3.1.1, Combination A match;

   b. "USB_MATCHES_ID" contains results of the Stage 1, Step 3.1.2, Combination B match;

   c. "USC_MATCHES_ID"contains results of the Stage 1, Step 3.1.2, Combination C match;

   d. "USD_MATCHES_ID" contains results of the Stage 1, Step 3.1.2, Combination D match;

   e. "USE1_MATCHES_ID" contains results of the Stage 1, Step 3.1.2, Combination E match;

   f. "USF_MATCHES_ID" contains results of the Stage 1, Step 3.1.2, Combination F match;

   g. "USG_MATCHES_ID" contains results of the Stage 2, Step 3.2.1, Combination G match;

   h. "USH_MATCHES_ID" contains results of the Stage 2, Step 3.2.2, Combination H match;

   i. "USI_MATCHES_ID" contains results of the Stage 2, Step 3.2.2, Combination I match;

   j. "USSSN_MATCHES_ID" contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;

k. "USK2_MATCHES_ID"contains results of the Stage 2, Step 3.2.3, Combination K to G match;

l. "USK_MATCHES_ID" contains results of the Stage 2, Step 3.2.3, Combination K match;

m. "USK3_MATCHES_ID" contains results of the Stage 2, Step 3.2.3, Combination K to L match;

n. "USL2_MATCHES_ID"contains results of the Stage 2, Step 3.2.4, Combination L to G match;

o. "USL_MATCHES_ID" contains results of the Stage 2, Step 3.2.4, Combination L match;

p. "USL3_MATCHES_ID"contains results of the Stage 2, Step 3.2.4, Combination L to K match;

q. "USF_NW_MATCHES_ID" contains results of the Stage 3, Step 3.3.1, Combination F match;

r. "USG_NW_MATCHES_ID" contains results of the Stage 3, Step 3.3.2, Combination G match;

s. "USI_NW_MATCHES_ID" contains results of the Stage 3, Step 3.3.2, Combination I match;

t. "USSSN_NW_MATCHES_ID" contains results of the Stage 3, Step 3.3.2, nine-digit social security number match;

u. "USK2_NW_MATCHES_ID" contains results of the Stage 3, Step 3.3.3, Combination K to G match;

v. "USK_NW_MATCHES_ID" contains results of the Stage 2, Step 3.3.3, Combination K match;

w. "USK3_NW_MATCHES_ID" contains results of the Stage 2, Step 3.3.3, Combination K to L match;

x. "USL2_NW_MATCHES_ID" contains results of the Stage 3, Step 3.3.4, Combination L to G match;

y. "USL_NW_MATCHES_ID" contains results of the Stage 3, Step 3.3.4, Combination L match;

z. "USL3_NW_MATCHES_ID" contains results of the Stage 3, Step 3.3.4, Combination L to K match;

aa. "TEXAS1_MATCHES_ID" contains results of the Stage 4, Step 3.4.1, Sweep 1 match;

bb. "TEXAS2_MATCHES_ID" contains results of the Stage 4, Step 3.4.2, Sweep 2 match;

cc. "TEXAS3_MATCHES_ID" contains results of the Stage 4, Step 3.4.3, Sweep 3 match; and

dd. "TEXAS4_MATCHES_ID" contains results of the Stage 4, Step 3.4.4, Sweep 4 match.

10. No other matching comparisons between the Texas data and data from the VBA were undertaken beyond those set forth above, in accordance with the steps provided in Ex. A.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on May 27, 2014.

Kenneth P. Smith

Exhibit A

***United States v. Texas*: Federal Agency Algorithm Summary**

This document summarizes the database matches that both the United States and all other plaintiffs and the State of Texas have requested from Federal Agencies as part of the *Veasey v. Perry*/*United States v. Texas* litigation (S.D. Tex).

The matching process proceeds in three parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, the United States' one-to-many matches and the State of Texas's many-to-many matches are conducted between databases.

PART I: DATABASE PREPARATION

Stage 1: Extraction of Available Data from Federal Identification & Disability Databases

Step 1.1.1: Extract complete name into separate first name, middle name, and last name fields.

Step 1.1.2: Extract date of birth.

Step 1.1.3: Extract gender.

Step 1.1.4: Extract residential address and mailing address.

Step 1.1.5: Extract social security number.

Step 1.1.6: Extract Texas driver license number (only if present in Federal database).

Stage 2: Separate Valid Identification and Disability Records

Step 1.2.1: Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

Exhibit A

Step 1.2.2: Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

Stage 3: Diagnostics

Step 1.3.1: Report the frequency of missing values for each field.

Step 1.3.2: Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.

Step 1.3.3: Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

Stage 4: Standardize Last Name

Step 1.4.1: Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.

Step 1.4.2: For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.

Step 1.4.3: Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.

Step 1.4.4: Code all missing values as blank fields.

Exhibit A

Stage 5: Standardize First Name and Middle Name

Step 1.5.1: Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.

Step 1.5.2: Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.[1]

Step 1.5.3: Code all missing values as blank fields.

Stage 6: Standardize Date of Birth

Step 1.6.1: Convert the date of birth to an eight-digit string of MMDDYYYY.

Step 1.6.2: Code all missing values as blank fields.

Stage 7: Standardize Gender

Step 1.7.1: Code gender as a string of 1 for females and 0 for males.

Step 1.7.2: Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.

Step 1.7.3: If missing values remain, code all missing values as blank fields.

[1] The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Stage 8: Standardize Address

Step 1.8.1: Convert the residential ZIP code to a string if it is stored as a numeric field.

Step 1.8.2: Where the residential address ZIP code is blank, populate that field with the value in the mailing address ZIP code field, if available.[2]

Step 1.8.3: Truncate the residential ZIP code field to the first five digits. *E.g.*, <77777-1234> becomes <77777>.

Step 1.8.4: Where the residential address field is blank, populate that field with the value in the mailing address field, if available.

Step 1.8.5: Where address field containing street address begins with a street number, isolate the street number. *E.g.*, <123 Main Street> becomes <123>.

Step 1.8.6: Where the address field begins with recognized strings indicating a mail box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444> becomes <444>.

Step 1.8.7: If missing values remain, code all missing values as blank fields.

Stage 9: Standardize Social Security Number

Step 1.9.1: Convert the social security number to a string if it is stored as a numeric field.

Step 1.9.2: Using full social security number, check for invalid SSNs. In the case of invalid SSNs, code as missing. *E.g.*, <123456789> becomes <>.

[2] For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Step 1.9.3: Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

Step 1.9.4: Code all missing values as blank fields.

PART II: DATABASE PREPARATION

Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps

Step 2.1.1: Create Combination A: First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.[3]

Step 2.1.2: Create Combination B: Last Name + Gender + DOB + Residential ZIP + Residential Street Number.

Step 2.1.3: Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.4: Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.5: Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.

Step 2.1.6: Create Combination F: First Name + Last Name + Gender + DOB.

Step 2.1.7: Create Combination M: Texas Driver License Number (where available).

[3] For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Exhibit A

Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps

Step 2.2.1: Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.[4]

Step 2.2.2: Create Combination H: SSN4 + Date of Birth + Residential ZIP.

Step 2.2.3: Create Combination I: SSN4 + First Name + Last Name + Date of Birth.

Step 2.2.4: Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.

Step 2.2.5: Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.

Step 2.2.6: Full Social Security Number.

Stage 3: Construct Identifiers Used Only For Texas's Many-to-Many Sweeps

Step 2.3.1: Create Combination for Texas's Sweep 1: SSN4 + Last Name + DOB.

Step 2.3.2: Create Combination for Texas's Sweep 3: First Name + Last Name + DOB

Step 2.3.3: Create Combination for Texas's Sweep 4: First Name + Middle Initial + Last Name + DOB

Note: Combinations for Texas's Sweeps 1 and 3 do not already exist as pre-made fields in the TEAM database extract but instead must be created from the underlying TEAM database fields, in addition to being constructed on the Federal database side. Texas's Sweep 4 is equivalent to

[4] Only for the State Department, create three further variations of Combination G created using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

the combination for the United States' Combination G. Texas's Sweep 2 is on full 9 social security number.[5]

Stage 4: Establish Identifier Uniqueness For Combinations A - L

Step 2.4.1: Generate a field that establishes the uniqueness of each identifier variable. For federal databases, for each combination A-L, generate a field that establishes uniqueness among only Texas records and a field that establishes uniqueness among nationwide records. *E.g.*, if only one record has the string <JEANSMITH01011950012377777> for Combination A, populate the uniqueness field for Combination A for that record as <1>. If four records have the string <JOHNSMITHA0101950> for Combination G, populate the uniqueness field for Combination G for each of those records as <2>, which indicates any number greater than one.

[5] For purposes of matching to U.S. Department of State Passport and Passport Card holder data, Texas's Sweep 3 and Sweep 4 are as follows: Sweep 3: Last Name + First Name (restricted to 32 characters) + DOB. Sweep 4: Last Name + First Name + Middle Initial (restricted to 32 characters) + DOB.

Exhibit A

PART III: MATCH DATABASES

Stage 1: United States' Primary One-to-Many Matching Sweeps

Step 3.1.1: For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identifier or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).

Step 3.1.2: Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identifier or disability database.

Step 3.1.3: Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identifier databases produced by the State of Texas.

Exhibit A

Stage 2: United States' Secondary One-to-Many Matching Sweeps

Step 3.2.1: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identifier or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.2.2: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number[6] in the TEAM database against the equivalent combination/field in the identifier or disability database.

Step 3.2.3: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identifier or disability database.

[6] The full social security number is not created as a separate "combination" as it is its own field stored within the TEAM database under the field name "ssn".

Exhibit A

Step 3.2.4: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identifier or disability database.

Stage 3: United States' Nationwide Federal Sweeps

Step 3.3.1: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identifier or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.3.2: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identifier or disability database.

Step 3.3.3: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination

K against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.

Step 3.3.4: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination L against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.[7]

Stage 4: Texas' Many-to-Many Nationwide Sweeps

Step 3.4.1 Regardless of whether the combination for Sweep 1 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

[7] Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last_fix + First_fix + Middle_name from the TEAM database;
- Combination G2 to DOB + Last_fix + First_fix from the TEAM database;
- Combination G2 to DOB + Last_fix + First word of First_name from the TEAM database;
- Combination G3 to DOB + Last_fix + First_fix + Middle_Initial from the TEAM database;
- Combination G3 to DOB + Last_fix + First word of First_name + Middle_Initial from the TEAM database; and
- Combination G3 to DOB + Last_fix + First word of First_name + First character of Second word of First_fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

Step 3.4.2 Regardless of whether full 9 social security number is unique in the TEAM database, for Sweep 2, match against the equivalent field in a nationwide search of the Federal database.

Step 3.4.3 Regardless of whether the combination for Sweep 3 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Step 3.4.4 Regardless of whether the combination for Sweep 4 (Combination G) is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Note: For each of the Texas many-to-many sweeps:

- Indicate <1> if any TEAM combination matches a single combination in the Federal database
- Indicate <2> if any TEAM combination matches more than one record in the Federal database.
- Indicate <0> if no match is achieved.

Examples:

- If there are two TEAM records that have identical versions of the combination for Sweep 1, and there is one record in the Federal database that matches on this combination, both of the underlying TEAM records will have a matching output of <1> for Sweep 1.

- If there are three TEAM records that have identical versions of the combination for Sweep 3, and there are five records in the Federal database that match on that combination, the three TEAM records will each have a matching output of <2> for Sweep 3.

Exhibit B

Veterans Benefits Administration

Step 1.3 Diagnostics on VBA Data

Total number of VBA records for all living Veterans rated as having 50% or higher disability, either service connected or non-service connected: 2,203,046

Blank field counts:
Last Name – 0
First Name – 0
Middle Name – 261,9448, or 11.89%
SSN – 4,211, or 0.191%
Date of Birth – 316, or 0.014%
Gender – 41,812, or 1.898%
Mailing Address – 1,639 or 0.074%
Residence Address – 1,735,958, or 78.798%

Counts of invalid SSNs:
111111111 – 0
222222222 – 0
333333333 – 0
444444444 – 0
555555555 – 0
666666666 – 0
777777777 – 0
888888888 – 0
999999999 – 0
000000000 – 0
012345678 – 0
123456789 – 0
987654321 – 0

Counts of potentially invalid dates of birth:
01-01-1901 – 5
02-02-1902 – 0
03-03-1903 – 0
04-04-1904 – 0
05-05-1905 – 0
06-06-1906 – 0
07-07-1907 – 1
08-08-1908 – 0
09-09-1909 – 0
10-10-1910 – 0
11-11-1911 – 1
12-12-1912 – 1

Exhibit H

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*, Plaintiffs, v. RICK PERRY, *et al.*, Defendants.	Civil Action No. 2:13-cv-193 (NGR)
UNITED STATES OF AMERICA, Plaintiff, TEXAS LEAGUE OF YOUNG VOTERS EDUCATION FUND, *et al.*, Plaintiff-Intervenors, TEXAS ASSOCIATION OF HISPANIC COUNTY JUDGES AND COUNTY COMMISSIONERS, *et al.*, Plaintiff-Intervenors, v. STATE OF TEXAS, *et al.*, Defendants.	Civil Action No. 2:13-cv-263 (NGR)

TEXAS STATE CONFERENCE OF NAACP BRANCHES, *et al.*, Plaintiffs, v. JOHN STEEN, *et al.*, Defendants.	Civil Action No. 2:13-cv-291 (NGR)
BELINDA ORTIZ, *et al.*, Plaintiffs, v. STATE OF TEXAS, *et al.*, Defendants	Civil Action No. 2:13-cv-348 (NGR)

DECLARATION OF MICHAEL MIMS

I, Michael Mims, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as Program Manager for HealthCare Identity Management for the Veterans Health Administration of the U.S. Department of Veterans Affairs. I have served in this capacity for April 2012. My duties include the maintenance and validation of the identity data stored within the Master Veteran Index.
2. I have personal knowledge of the information contained in this declaration based upon my work for the Veterans Health Administration of the United States Department of Veterans Affairs ("VHA"), and my personal involvement in supervision of the database comparison described below.

3. On March 11, 2014, my agency received an encrypted hard drive containing data regarding Texas registered voters from the United States Department of Justice in the form of a 7 gigabyte .csv file containing 13,564,420 records for Texas registered voters. This file was corrupted in the process of loading the data and so an identical dataset was delivered to VHA on March 26, 2014. VHA was able to load all 13,564,420 records.

4. The Texas data that VHA received contained the following fields:

- Vuid
- last_name
- first_name
- middle_name
- ssn
- ssn4_str
- gender_code
- official_id
- dl_id_str
- date_of_birth
- voter_status_code
- block_number
- street_name
- city
- state
- zip_code
- zip_str
- mail_adrs_1
- last_fix
- multi_lastname1
- multi_lastname2
- first_fix
- middle_initial
- fem_str
- dob_str
- zip5_res_alt
- str_num_res_alt
- AGDN
- AGDN_TAG
- AGDN_unique
- AGDNlast
- AGDNlast_TAG
- AGDNlast_unique
- AGD
- AGD_TAG
- AGD_unique
- ADN
- ADN_TAG
- ADN_unique
- AGN
- AGN_TAG
- AGN_unique
- GDN
- GDN_TAG
- GDN_unique
- DLN
- DLN_TAG
- DLN_unique
- unique_count
- FMLD
- FMLD_TAG
- FMLD_unique
- SDZ
- SDZ_TAG
- SDZ_unique
- SND
- SND_TAG
- SND_unique
- FML1D
- FML1D_TAG
- FML1D_unique
- FML2D

- FML2D_TAG
- FML2D_unique
- unique_count2
- var_fill
- K2
- K3
- L2
- L3

5. From the fields listed above, VHA loaded the following variables that were specifically used in completing the matching process:

- vuid
- GDN
- FMLD
- FML1D
- FML2D
- AGDN
- AGDNlast
- AGD
- ADN
- AGN
- SSN
- ssn4_str
- last_fix
- first_fix
- middle_initial
- dob_str
- SDZ
- SND
- var_fill
- K2
- K3
- L2
- L3

6. VHA executed the data preparation and comparison steps, attached as Ex. A, that were provided by the United States Department of Justice.

7. Through this process, VHA attempted to match particular combinations of identifying information for Texas registered voters (*e.g.*, first and last name, gender, and date of birth) with the same or related combinations of identifying information with respect to holders of a Veteran Identification Card. For each particular combination, VHA identified instances where the identifying information for a Texas registered voter matched with the same or related combinations of identifying information for one or more holders of a Veteran Identification Card. The frequencies of any missing values in the underlying VHA dataset are provided in Ex. B.

8. For each sweep through the relevant VHA dataset, we appended a column to the Texas data that VHA loaded to indicate, on a record-by-record basis, the output of the database

comparisons on all of the requested combinations where there was one or more matches. In total, out of the 13,564,420 records available to be matched, VHA found matches on one or more combinations for 299,320 voter records.

9. On April 30, 2014, VHA transferred responsive data to the Department of Justice, in the form of a 117 megabyte text (.txt) file, with vertical bar delimiters. This file contained the match results for each of the 299,320 records where VHA found one or more matches. The voter file records where no matches were found were not returned.

10. The results of each sweep, as described in Ex. A, can be found in the following columns of the data VHA returned to the Department of Justice:

    a. "VHA_USA" contains results of the Stage 1, Step 3.1.1, Combination A match;

    b. "VHA_USB" contains results of the Stage 1, Step 3.1.2, Combination B match;

    c. "VHA_USC"contains results of the Stage 1, Step 3.1.2, Combination C match;

    d. "VHA_USD" contains results of the Stage 1, Step 3.1.2, Combination D match;

    e. "VHA_USE" contains results of the Stage 1, Step 3.1.2, Combination E match;

    f. "VHA_USF" contains results of the Stage 1, Step 3.1.2, Combination F match;

    g. "VHA_USG" contains results of the Stage 2, Step 3.2.1, Combination G match;

    h. "VHA_USH" contains results of the Stage 2, Step 3.2.2, Combination H match;

    i. "VHA_USI" contains results of the Stage 2, Step 3.2.2, Combination I match;

    j. "VHA_USSSN" contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;

    k. "VHA_USK2"contains results of the Stage 2, Step 3.2.3, Combination K to G match;

    l. "VHA_USK" contains results of the Stage 2, Step 3.2.3, Combination K match;

    m. "VHA_USK3" contains results of the Stage 2, Step 3.2.3, Combination K to L match;

n. "VHA_USL2"contains results of the Stage 2, Step 3.2.4, Combination L to G match;

o. "VHA_USL" contains results of the Stage 2, Step 3.2.4, Combination L match;

p. "VHA_USL3"contains results of the Stage 2, Step 3.2.4, Combination L to K match;

q. "VHA_USF_NW" contains results of the Stage 3, Step 3.3.1, Combination F match;

r. "VHA_USG_NW" contains results of the Stage 3, Step 3.3.2, Combination G match;

s. "VHA_USI_NW" contains results of the Stage 3, Step 3.3.2, Combination I match;

t. "VHA_USSSN_NW" contains results of the Stage 3, Step 3.3.2, nine-digit social security number match;

u. "VHA_USK2_NW" contains results of the Stage 3, Step 3.3.3, Combination K to G match;

v. "VHA_USK_NW" contains results of the Stage 2, Step 3.3.3, Combination K match;

w. "VHA_USK3_NW" contains results of the Stage 2, Step 3.3.3, Combination K to L match;

x. "VHA_USL2_NW" contains results of the Stage 3, Step 3.3.4, Combination L to G match;

y. "VHA_USL_NW" contains results of the Stage 3, Step 3.3.4, Combination L match;

z. "VHA_USL3_NW" contains results of the Stage 3, Step 3.3.4, Combination L to K match;

aa. "texas1" contains results of the Stage 4, Step 3.4.1, Sweep 1 match;

bb. "texas2" contains results of the Stage 4, Step 3.4.2, Sweep 2 match;

cc. "texas3" contains results of the Stage 4, Step 3.4.3, Sweep 3 match; and

dd. "texas4" contains results of the Stage 4, Step 3.4.4, Sweep 4 match.

11. No other matching comparisons between the Texas data and data from VHA were undertaken beyond those set forth in Ex. A.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on May 27, 2014.

Exhibit A

***United States v. Texas*: Federal Agency Algorithm Summary**

This document summarizes the database matches that both the United States and all other plaintiffs and the State of Texas have requested from Federal Agencies as part of the *Veasey v. Perry*/*United States v. Texas* litigation (S.D. Tex).

The matching process proceeds in three parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, the United States' one-to-many matches and the State of Texas's many-to-many matches are conducted between databases.

PART I: DATABASE PREPARATION

Stage 1: Extraction of Available Data from Federal Identification & Disability Databases

Step 1.1.1: Extract complete name into separate first name, middle name, and last name fields.

Step 1.1.2: Extract date of birth.

Step 1.1.3: Extract gender.

Step 1.1.4: Extract residential address and mailing address.

Step 1.1.5: Extract social security number.

Step 1.1.6: Extract Texas driver license number (only if present in Federal database).

Stage 2: Separate Valid Identification and Disability Records

Step 1.2.1: Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

Step 1.2.2: Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

Stage 3: Diagnostics

Step 1.3.1: Report the frequency of missing values for each field.

Step 1.3.2: Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.

Step 1.3.3: Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

Stage 4: Standardize Last Name

Step 1.4.1: Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.

Step 1.4.2: For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.

Step 1.4.3: Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.

Step 1.4.4: Code all missing values as blank fields.

Exhibit A

Stage 5: Standardize First Name and Middle Name

Step 1.5.1: Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.

Step 1.5.2: Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.[1]

Step 1.5.3: Code all missing values as blank fields.

Stage 6: Standardize Date of Birth

Step 1.6.1: Convert the date of birth to an eight-digit string of MMDDYYYY.

Step 1.6.2: Code all missing values as blank fields.

Stage 7: Standardize Gender

Step 1.7.1: Code gender as a string of 1 for females and 0 for males.

Step 1.7.2: Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.

Step 1.7.3: If missing values remain, code all missing values as blank fields.

[1] The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Exhibit A

Stage 8: Standardize Address

Step 1.8.1: Convert the residential ZIP code to a string if it is stored as a numeric field.

Step 1.8.2: Where the residential address ZIP code is blank, populate that field with the value in the mailing address ZIP code field, if available.[2]

Step 1.8.3: Truncate the residential ZIP code field to the first five digits. *E.g.*, <77777-1234> becomes <77777>.

Step 1.8.4: Where the residential address field is blank, populate that field with the value in the mailing address field, if available.

Step 1.8.5: Where address field containing street address begins with a street number, isolate the street number. *E.g.*, <123 Main Street> becomes <123>.

Step 1.8.6: Where the address field begins with recognized strings indicating a mail box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444> becomes <444>.

Step 1.8.7: If missing values remain, code all missing values as blank fields.

Stage 9: Standardize Social Security Number

Step 1.9.1: Convert the social security number to a string if it is stored as a numeric field.

Step 1.9.2: Using full social security number, check for invalid SSNs. In the case of invalid SSNs, code as missing. *E.g.*, <123456789> becomes <>.

[2] For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Step 1.9.3: Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

Step 1.9.4: Code all missing values as blank fields.

PART II: DATABASE PREPARATION

Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps

Step 2.1.1: Create Combination A: First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.[3]

Step 2.1.2: Create Combination B: Last Name + Gender + DOB + Residential ZIP + Residential Street Number.

Step 2.1.3: Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.4: Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.5: Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.

Step 2.1.6: Create Combination F: First Name + Last Name + Gender + DOB.

Step 2.1.7: Create Combination M: Texas Driver License Number (where available).

[3] For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Exhibit A

Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps

Step 2.2.1: Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.[4]

Step 2.2.2: Create Combination H: SSN4 + Date of Birth + Residential ZIP.

Step 2.2.3: Create Combination I: SSN4 + First Name + Last Name + Date of Birth.

Step 2.2.4: Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.

Step 2.2.5: Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.

Step 2.2.6: Full Social Security Number.

Stage 3: Construct Identifiers Used Only For Texas's Many-to-Many Sweeps

Step 2.3.1: Create Combination for Texas's Sweep 1: SSN4 + Last Name + DOB.

Step 2.3.2: Create Combination for Texas's Sweep 3: First Name + Last Name + DOB

Step 2.3.3: Create Combination for Texas's Sweep 4: First Name + Middle Initial + Last Name + DOB

Note: Combinations for Texas's Sweeps 1 and 3 do not already exist as pre-made fields in the TEAM database extract but instead must be created from the underlying TEAM database fields, in addition to being constructed on the Federal database side. Texas's Sweep 4 is equivalent to

[4] Only for the State Department, create three further variations of Combination G created using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

the combination for the United States' Combination G. Texas's Sweep 2 is on full 9 social security number.[5]

Stage 4: Establish Identifier Uniqueness For Combinations A - L

Step 2.4.1: Generate a field that establishes the uniqueness of each identifier variable. For federal databases, for each combination A-L, generate a field that establishes uniqueness among only Texas records and a field that establishes uniqueness among nationwide records. *E.g.*, if only one record has the string <JEANSMITH01011950012377777> for Combination A, populate the uniqueness field for Combination A for that record as <1>. If four records have the string <JOHNSMITHA0101950> for Combination G, populate the uniqueness field for Combination G for each of those records as <2>, which indicates any number greater than one.

[5] For purposes of matching to U.S. Department of State Passport and Passport Card holder data, Texas's Sweep 3 and Sweep 4 are as follows: Sweep 3: Last Name + First Name (restricted to 32 characters) + DOB. Sweep 4: Last Name + First Name + Middle Initial (restricted to 32 characters) + DOB.

Exhibit A

PART III: MATCH DATABASES

Stage 1: United States' Primary One-to-Many Matching Sweeps

Step 3.1.1: For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identifier or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).

Step 3.1.2: Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identifier or disability database.

Step 3.1.3: Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identifier databases produced by the State of Texas.

Exhibit A

Stage 2: United States' Secondary One-to-Many Matching Sweeps

Step 3.2.1: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identifier or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.2.2: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number[6] in the TEAM database against the equivalent combination/field in the identifier or disability database.

Step 3.2.3: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identifier or disability database.

[6] The full social security number is not created as a separate "combination" as it is its own field stored within the TEAM database under the field name "ssn".

Exhibit A

Step 3.2.4: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identifier or disability database.

Stage 3: United States' Nationwide Federal Sweeps

Step 3.3.1: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identifier or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.3.2: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identifier or disability database.

Step 3.3.3: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination

K against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.

Step 3.3.4: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination L against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.[7]

Stage 4: Texas' Many-to-Many Nationwide Sweeps

Step 3.4.1 Regardless of whether the combination for Sweep 1 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

[7] Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last_fix + First_fix + Middle_name from the TEAM database;
- Combination G2 to DOB + Last_fix + First_fix from the TEAM database;
- Combination G2 to DOB + Last_fix + First word of First_name from the TEAM database;
- Combination G3 to DOB + Last_fix + First_fix + Middle_Initial from the TEAM database;
- Combination G3 to DOB + Last_fix + First word of First_name + Middle_Initial from the TEAM database; and
- Combination G3 to DOB + Last_fix + First word of First_name + First character of Second word of First_fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

Step 3.4.2 Regardless of whether full 9 social security number is unique in the TEAM database, for Sweep 2, match against the equivalent field in a nationwide search of the Federal database.

Step 3.4.3 Regardless of whether the combination for Sweep 3 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Step 3.4.4 Regardless of whether the combination for Sweep 4 (Combination G) is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Note: For each of the Texas many-to-many sweeps:

- Indicate <1> if any TEAM combination matches a single combination in the Federal database
- Indicate <2> if any TEAM combination matches more than one record in the Federal database.
- Indicate <0> if no match is achieved.

Examples:

- If there are two TEAM records that have identical versions of the combination for Sweep 1, and there is one record in the Federal database that matches on this combination, both of the underlying TEAM records will have a matching output of <1> for Sweep 1.

- If there are three TEAM records that have identical versions of the combination for Sweep 3, and there are five records in the Federal database that match on that combination, the three TEAM records will each have a matching output of <2> for Sweep 3.

Exhibit B

Step 1.3 Diagnostics on the VHA Data

There were 5,020,207 records extracted from the VHA data for Veterans with a Veteran Identification Card (VIC) on 2/28/2014. The number of Veterans with a VIC and a Texas residential address was 389,383.

Name

	Number of Records	Percent of Records
Last Name Missing	0	0.00%
First Name Missing	31	0.00%
First Name Only One Character	11,881	0.24%
Middle Initial Missing	661,201	13.17%

Social Security Number

	Number of Records	Percent of Records
SSN Missing	2,481	0.05%
Invalid SSN (begins with '00000')	27	0.00%

Gender

Gender was 100% populated.

Date of Birth

The possibly invalid dates of birth below appear to be default values used in the date of birth (e.g., first day of month, 15th of month, first day of year). This is rare.

	Number of Records	Percent of Records
Incomplete date or outside 1901-1996	77	0.00%
Possibly invalid (01-01-01, etc.)	21	0.00%

Zip Code

	Number of Records	Percent of Records
Zip missing (Most of these were cases with a missing residential address)	132,103	2.63%

Exhibit B

Mailing Zip

There were 5,301 veterans with a mailing address.

	Number of Records	Percent of Records
Mailing Zip missing	20	0.38%

Exhibit I

STEPHEN DANIEL ANSOLABEHERE

Department of Government
Harvard University
1737 Cambridge Street
Cambridge, MA 02138
sda@gov.harvard.edu

EDUCATION

Harvard University	Ph.D., Political Science	1989
University of Minnesota	B.A., Political Science	1984
	B.S., Economics	

PROFESSIONAL EXPERIENCE

ACADEMIC POSITIONS

2008-present	Professor, Department of Government, Harvard University
1998-2009	Elting Morison Professor, Department of Political Science, MIT (Associate Head, 2001-2005)
1995-1998	Associate Professor, Department of Political Science, MIT
1993-1994	National Fellow, The Hoover Institution
1989-1993	Assistant Professor, Department of Political Science, University of California, Los Angeles

FELLOWSHIPS AND HONORS

American Academy of Arts and Sciences	2007
Carnegie Scholar	2000-02
Goldsmith Book Prize for *Going Negative*	1996
National Fellow, The Hoover Institution	1993-94
Harry S. Truman Fellowship	1982-86

PUBLICATIONS

Books

2014 *Cheap and Clean: How Americans Think About Energy in the Age of Global Warming*. (with David Konisky). MIT Press.

2011 *American Government*, 12th edition, W.W. Norton. With Benjamin Ginsberg and Kenneth Shepsle

2008 *The End of Inequality: One Person, One Vote and the Transformation of American Politics*. W. W. Norton.

1996 *Going Negative: How Political Advertising Divides and Shrinks the American Electorate* (with Shanto Iyengar). The Free Press.

1993 *Media Game: American Politics in the Television Age* (with Roy Behr and Shanto Iyengar). Macmillan.

Recent Articles in Refereed Journals

Forthcoming "Mecro-economic Voting: Local Information and Micro-Perceptions of the Macro-Economy" (With Marc Meredith and Erik Snowberg), *Economics and Politics* (accepted)

2013 "Race, Gender, Age, and Voting" *Politics and Governance*, vol. 1, issue 2. (with Eitan Hersh)
http://www.librelloph.com/politicsandgovernance/article/view/PaG-1.2.132

2013 "Regional Differences in Racially Polarized Voting: Implications for the Constitutionality of Section 5 of the Voting Rights Act" (with Nathaniel Persily and Charles Stewart) 126 *Harvard Law Review* F 205 (2013) http://www.harvardlawreview.org/issues/126/april13/forum_1005.php

2013 "Cooperative Survey Research" *Annual Review of Political Science* (with Douglas Rivers)

2013 "Social Sciences and the Alternative Energy Future" *Daedalus* (with Bob Fri)

2013 "The Effects of Redistricting on Incumbents," *Election Law Journal* (with James Snyder)

2013 "Does Survey Mode Still Matter?" *Political Analysis* (with Brian Schaffner)

2012 "Asking About Numbers: How and Why" *Political Analysis* (with Erik Snowberg and Marc Meredith). doi:10.1093/pan/mps031

2012 "Movers, Stayers, and Registration" *Quarterly Journal of Political Science* (with Eitan Hersh and Ken Shepsle)

2012 "Validation: What Big Data Reveals About Survey Misreporting and the Real Electorate" *Political Analysis* (with Eitan Hersh)

2012 "Arizona Free Enterprise v. Bennett and the Problem of Campaign Finance" *Supreme Court Review* 2011(1):39-79

2012 "The American Public's Energy Choice" *Daedalus* (with David Konisky)

2012 "Challenges for Technology Change" *Daedalus* (with Robert Fri)

2011 "When Parties Are Not Teams: Party positions in single-member district and proportional representation systems" *Economic Theory* 49 (March) DOI: 10.1007/s00199-011-0610-1

2011 "Profiling Originalism" *Columbia Law Review* (with Jamal Greene and Nathaniel Persily).

2010 "Partisanship, Public Opinion, and Redistricting" *Election Law Journal* (with Joshua Fougere and Nathaniel Persily).

2010 "Primary Elections and Party Polarization" *Quarterly Journal of Political Science* (with Shigeo Hirano, James Snyder, and Mark Hansen)

2010 "Constituents' Responses to Congressional Roll Call Voting," *American Journal of Political Science* (with Phil Jones)

2010 "Race, Region, and Stephen Ansolabehere, Nathaniel Persily, and Charles H. Stewart III, "Race, Region, and Vote Choice in the 2008 Election: Implications for the Future of the Voting Rights Act" *Harvard Law Review* April, 2010.

2010 "Residential Mobility and the Cell Only Population," *Public Opinion Quarterly* (with Brian Schaffner)

2009 "Explaining Attitudes Toward Power Plant Location," *Public Opinion Quarterly* (with David Konisky)

2009 "Public risk perspectives on the geologic storage of carbon dioxide," *International Journal of Greenhouse Gas Control* (with Gregory Singleton and Howard Herzog) 3(1): 100-107.

2008 "A Spatial Model of the Relationship Between Seats and Votes" (with William Leblanc) *Mathematical and Computer Modeling* (November).

2008 "The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting" (with Jonathan Rodden and James M. Snyder, Jr.) *American Political Science Review* (May).

2008 "Access versus Integrity in Voter Identification Requirements." *New York University Annual Survey of American Law,* vol 63.

2008 "Voter Fraud in the Eye of the Beholder" (with Nathaniel Persily) *Harvard Law Review* (May)

2007 "Incumbency Advantages in U. S. Primary Elections," (with John Mark Hansen, Shigeo Hirano, and James M. Snyder, Jr.) *Electoral Studies* (September)

2007 "Television and the Incumbency Advantage" (with Erik C. Snowberg and James M. Snyder, Jr). *Legislative Studies Quarterly*.

2006 "The Political Orientation of Newspaper Endorsements" (with Rebecca Lessem and James M. Snyder, Jr.). *Quarterly Journal of Political Science* vol. 1, issue 3.

2006 "Voting Cues and the Incumbency Advantage: A Critical Test" (with Shigeo Hirano, James M. Snyder, Jr., and Michiko Ueda) *Quarterly Journal of Political Science* vol. 1, issue 2.

2006 "American Exceptionalism? Similarities and Differences in National Attitudes Toward Energy Policies and Global Warming" (with David Reiner, Howard Herzog, K. Itaoka, M. Odenberger, and Fillip Johanssen) *Environmental Science and Technology* (February 22, 2006), http://pubs3.acs.org/acs/journals/doilookup?in_doi=10.1021/es052010b

2006 "Purple America" (with Jonathan Rodden and James M. Snyder, Jr.) *Journal of Economic Perspectives* (Winter).

2005 "Did the Introduction of Voter Registration Decrease Turnout?" (with David Konisky). *Political Analysis*.

2005 "Statistical Bias in Newspaper Reporting: The Case of Campaign Finance" *Public Opinion Quarterly* (with James M. Snyder, Jr., and Erik Snowberg).

2005 "Studying Elections" *Policy Studies Journal* (with Charles H. Stewart III and R. Michael Alvarez).

2005 "Legislative Bargaining under Weighted Voting" *American Economic Review* (with James M. Snyder, Jr., and Michael Ting)

2005 "Voting Weights and Formateur Advantages in Coalition Formation: Evidence from Parliamentary Coalitions, 1946 to 2002" (with James M. Snyder, Jr., Aaron B. Strauss, and Michael M. Ting) *American Journal of Political Science*.

2005 "Reapportionment and Party Realignment in the American States" *Pennsylvania Law Review* (with James M. Snyder, Jr.)

2004 "Residual Votes Attributable to Voting Technologies" (with Charles Stewart) *Journal of Politics*

2004 "Using Term Limits to Estimate Incumbency Advantages When Office Holders Retire Strategically" (with James M. Snyder, Jr.). *Legislative Studies Quarterly* vol. 29, November 2004, pages 487-516.

2004 "Did Firms Profit From Soft Money?" (with James M. Snyder, Jr., and Michiko Ueda) *Election Law Journal* vol. 3, April 2004.

2003 "Bargaining in Bicameral Legislatures" (with James M. Snyder, Jr. and Mike Ting) *American Political Science Review,* August, 2003.

2003 "Why Is There So Little Money in U.S. Politics?" (with James M. Snyder, Jr.) *Journal of Economic Perspectives*, Winter, 2003.

2002 "Equal Votes, Equal Money: Court-Ordered Redistricting and the Public Spending in the American States" (with Alan Gerber and James M. Snyder, Jr.) *American Political Science Review*, December, 2002.
Paper awarded the Heinz Eulau award for the best paper in the American Political Science Review.

2002 "Are PAC Contributions and Lobbying Linked?" (with James M. Snyder, Jr. and Micky Tripathi) *Business and Politics* 4, no. 2.

2002 "The Incumbency Advantage in U.S. Elections: An Analysis of State and Federal Offices, 1942-2000" (with James Snyder) *Election Law Journal,* 1, no. 3.

2001 "Voting Machines, Race, and Equal Protection." *Election Law Journal*, vol. 1, no. 1

2001 "Models, assumptions, and model checking in ecological regressions" (with Andrew Gelman, David Park, Phillip Price, and Larraine Minnite) *Journal of the Royal Statistical Society,* series A, 164: 101-118.

2001 "The Effects of Party and Preferences on Congressional Roll Call Voting." (with James Snyder and Charles Stewart) *Legislative Studies Quarterly* (forthcoming).
Paper awarded the *Jewell-Lowenberg Award* for the best paper published on legislative politics in 2001. Paper awarded the *Jack Walker Award* for the best paper published on party politics in 2001.

2001 "Candidate Positions in Congressional Elections," (with James Snyder and Charles Stewart). *American Journal of Political Science* 45 (November).

2000 "Old Voters, New Voters, and the Personal Vote," (with James Snyder and Charles Stewart) *American Journal of Political Science* 44 (February).

2000 "Soft Money, Hard Money, Strong Parties," (with James Snyder) *Columbia Law Review* 100 (April):598 - 619.

2000 "Campaign War Chests and Congressional Elections," (with James Snyder) *Business and Politics*. 2 (April): 9-34.

1999 "Replicating Experiments Using Surveys and Aggregate Data: The Case of Negative Advertising." (with Shanto Iyengar and Adam Simon) *American Political Science Review* 93 (December).

1999 "Valence Politics and Equilibrium in Spatial Models," (with James Snyder), *Public Choice.*

1999 "Money and Institutional Power," (with James Snyder), *Texas Law Review* 77 (June, 1999): 1673-1704.

1997 "Incumbency Advantage and the Persistence of Legislative Majorities," (with Alan Gerber), *Legislative Studies Quarterly* 22 (May 1997).

1996 "The Effects of Ballot Access Rules on U.S. House Elections," (with Alan Gerber), *Legislative Studies Quarterly* 21 (May 1996).

1994 "Riding the Wave and Issue Ownership: The Importance of Issues in Political Advertising and News," (with Shanto Iyengar) *Public Opinion Quarterly* 58:

335-357.

1994 "Horseshoes and Horseraces: Experimental Evidence of the Effects of Polls on Campaigns," (with Shanto Iyengar) *Political Communications* 11/4 (October-December): 413-429.

1994 "Does Attack Advertising Demobilize the Electorate?" (with Shanto Iyengar), *American Political Science Review* 89 (December).

1994 "The Mismeasure of Campaign Spending: Evidence from the 1990 U.S. House Elections," (with Alan Gerber) *Journal of Politics* 56 (September).

1993 "Poll Faulting," (with Thomas R. Belin) *Chance* 6 (Winter): 22-28.

1991 "The Vanishing Marginals and Electoral Responsiveness," (with David Brady and Morris Fiorina) *British Journal of Political Science* 22 (November): 21-38.

1991 "Mass Media and Elections: An Overview," (with Roy Behr and Shanto Iyengar) *American Politics Quarterly* 19/1 (January): 109-139.

1990 "The Limits of Unraveling in Interest Groups," *Rationality and Society* 2: 394-400.

1990 "Measuring the Consequences of Delegate Selection Rules in Presidential Nominations," (with Gary King) *Journal of Politics* 52: 609-621.

1989 "The Nature of Utility Functions in Mass Publics," (with Henry Brady) *American Political Science Review* 83: 143-164.

Special Reports and Policy Studies

2010 *The Future of Nuclear Power*, Revised.

2006 *The Future of Coal.* MIT Press. Continued reliance on coal as a primary power source will lead to very high concentrations of carbon dioxide in the atmosphere, resulting in global warming. This cross-disciplinary study – drawing on faculty from Physics, Economics, Chemistry, Nuclear Engineering, and Political Science – develop a road map for technology research and development policy in order to address the challenges of carbon emissions from expanding use of coal for electricity and heating throughout the world.

2003 *The Future of Nuclear Power*. MIT Press. This cross-disciplinary study – drawing on faculty from Physics, Economics, Chemistry, Nuclear Engineering,

and Political Science – examines the what contribution nuclear power can make to meet growing electricity demand, especially in a world with increasing carbon dioxide emissions from fossil fuel power plants.

2002 "Election Day Registration." A report prepared for DEMOS. This report analyzes the possible effects of Proposition 52 in California based on the experiences of 6 states with election day registration.

2001 *Voting: What Is, What Could Be.* A report of the Caltech/MIT Voting Technology Project. This report examines the voting system, especially technologies for casting and counting votes, registration systems, and polling place operations, in the United States. It was widely used by state and national governments in formulating election reforms following the 2000 election.

2001 "An Assessment of the Reliability of Voting Technologies." A report of the Caltech/MIT Voting Technology Project. This report provided the first nationwide assessment of voting equipment performance in the United States. It was prepared for the Governor's Select Task Force on Election Reform in Florida.

Chapters in Edited Volumes

2012 "Using Recounts to Measure the Accuracy of Vote Tabulations: Evidence from New Hampshire Elections, 1946-2002" in Confirming Elections, R. Michael Alvarez, Lonna Atkeson, and Thad Hall, eds. New York: Palgrave, Macmillan.

2010 "Dyadic Representation" in Oxford Handbook on Congress, Eric Schickler, ed., Oxford University Press.

2008 "Voting Technology and Election Law" in *America Votes!,* Benjamin Griffith, editor, Washington, DC: American Bar Association.

2007 "What Did the Direct Primary Do to Party Loyalty in Congress" (with Shigeo Hirano and James M. Snyder Jr.) in *Process, Party and Policy Making: Further New Perspectives on the History of Congress*, David Brady and Matthew D. McCubbins (eds.), Stanford University Press, 2007.

2007 "Election Administration and Voting Rights" in *Renewal of the Voting Rights Act*, David Epstein and Sharyn O'Hallaran, eds. Russell Sage Foundation.

2006 "The Decline of Competition in Primary Elections," (with John Mark Hansen, Shigeo Hirano, and James M. Snyder, Jr.) *The Marketplace of Democracy*, Michael P. McDonald and John Samples, eds. Washington, DC: Brookings.

2005 "Voters, Candidates and Parties" in *Handbook of Political Economy*, Barry

Weingast and Donald Wittman, eds. New York: Oxford University Press.

2003 "Baker v. Carr in Context, 1946 – 1964" (with Samuel Isaacharoff) in *Constitutional Cases in Contex*t, Michael Dorf, editor. New York: Foundation Press.

2002 "Corruption and the Growth of Campaign Spending"(with Alan Gerber and James Snyder). *A User's Guide to Campaign Finance*, Jerry Lubenow, editor. Rowman and Littlefield.

2001 "The Paradox of Minimal Effects," in Henry Brady and Richard Johnston, eds., *Do Campaigns Matter*? University of Michigan Press.

2001 "Campaigns as Experiments," in Henry Brady and Richard Johnson, eds., Do *Campaigns Matter*? University of Michigan Press.

2000 "Money and Office," (with James Snyder) in David Brady and John Cogan, eds., *Congressional Elections: Continuity and Change*. Stanford University Press.

1996 "The Science of Political Advertising," (with Shanto Iyengar) in *Political Persuasion and Attitude Change*, Richard Brody, Diana Mutz, and Paul Sniderman, eds. Ann Arbor, MI: University of Michigan Press.

1995 "Evolving Perspectives on the Effects of Campaign Communication," in Philo Warburn, ed., *Research in Political Sociology*, vol. 7, JAI.

1995 "The Effectiveness of Campaign Advertising: It's All in the Context," (with Shanto Iyengar) in *Campaigns and Elections American Style*, Candice Nelson and James A. Thurber, eds. Westview Press.

1993 "Information and Electoral Attitudes: A Case of Judgment Under Uncertainty," (with Shanto Iyengar), in *Explorations in Political Psychology*, Shanto Iyengar and William McGuire, eds. Durham: Duke University Press.

Working Papers

2009 "Sociotropic Voting and the Media" (with Marc Meredith and Erik Snowberg), American National Election Study Pilot Study Reports, John Aldrich editor.

2007 "Public Attitudes Toward America's Energy Options: Report of the 2007 MIT Energy Survey" CEEPR Working Paper 07-002 and CANES working paper.

2006 "Constituents' Policy Perceptions and Approval of Members' of Congress" CCES Working Paper 06-01 (with Phil Jones).

2004 "Using Recounts to Measure the Accuracy of Vote Tabulations: Evidence from New Hampshire Elections, 1946 to 2002" (with Andrew Reeves).

2002 "Evidence of Virtual Representation: Reapportionment in California," (with Ruimin He and James M. Snyder).

1999 "Why did a majority of Californians vote to lower their own power?" (with James Snyder and Jonathan Woon). Paper presented at the annual meeting of the American Political Science Association, Atlanta, GA, September, 1999. Paper received the award for the best paper on Representation at the 1999 Annual Meeting of the APSA.

1999 "Has Television Increased the Cost of Campaigns?" (with Alan Gerber and James Snyder).

1996 "Money, Elections, and Candidate Quality," (with James Snyder).

1996 "Party Platform Choice - Single- Member District and Party-List Systems,"(with James Snyder).

1995 "Messages Forgotten" (with Shanto Iyengar).

1994 "Consumer Contributors and the Returns to Fundraising: A Microeconomic Analysis," (with Alan Gerber), presented at the Annual Meeting of the American Political Science Association, September.

1992 "Biases in Ecological Regression," (with R. Douglas Rivers) August, (revised February 1994). Presented at the Midwest Political Science Association Meetings, April 1994, Chicago, IL.

1992 "Using Aggregate Data to Correct Nonresponse and Misreporting in Surveys" (with R. Douglas Rivers). Presented at the annual meeting of the Political Methodology Group, Cambridge, Massachusetts, July.

1991 "The Electoral Effects of Issues and Attacks in Campaign Advertising" (with Shanto Iyengar). Presented at the Annual Meeting of the American Political Science Association, Washington, DC.

1991 "Television Advertising as Campaign Strategy: Some Experimental Evidence" (with Shanto Iyengar). Presented at the Annual Meeting of the American Association for Public Opinion Research, Phoenix.

1991 "Why Candidates Attack: Effects of Televised Advertising in the 1990 California Gubernatorial Campaign," (with Shanto Iyengar). Presented at the Annual

Meeting of the Western Political Science Association, Seattle, March.

1990 "Winning is Easy, But It Sure Ain't Cheap." Working Paper #90-4, Center for the American Politics and Public Policy, UCLA. Presented at the Political Science Departments at Rochester University and the University of Chicago.

Research Grants

1989-1990 Markle Foundation. "A Study of the Effects of Advertising in the 1990 California Gubernatorial Campaign." Amount: $50,000

1991-1993 Markle Foundation. "An Experimental Study of the Effects of Campaign Advertising." Amount: $150,000

1991-1993 NSF. "An Experimental Study of the Effects of Advertising in the 1992 California Senate Electoral." Amount: $100,000

1994-1995 MIT Provost Fund. "Money in Elections: A Study of the Effects of Money on Electoral Competition." Amount: $40,000

1996-1997 National Science Foundation. "Campaign Finance and Political Representation." Amount: $50,000

1997 National Science Foundation. "Party Platforms: A Theoretical Investigation of Party Competition Through Platform Choice." Amount: $40,000

1997-1998 National Science Foundation. "The Legislative Connection in Congressional Campaign Finance. Amount: $150,000

1999-2000 MIT Provost Fund. "Districting and Representation." Amount: $20,000.

1999-2002 Sloan Foundation. "Congressional Staff Seminar." Amount: $156,000.

2000-2001 Carnegie Corporation. "The Caltech/MIT Voting Technology Project." Amount: $253,000.

2001-2002 Carnegie Corporation. "Dissemination of Voting Technology Information." Amount: $200,000.

2003-2005 National Science Foundation. "State Elections Data Project." Amount: $256,000.

2003-2004 Carnegie Corporation. "Internet Voting." Amount: $279,000.

2003-2005 Knight Foundation. "Accessibility and Security of Voting Systems." Amount: $450,000.

2006-2008 National Science Foundation, "Primary Election Data Project"

2008-2009 Pew/JEHT. "Measuring Voting Problems in Primary Elections, A National Survey." Amount: $300,000

2008-2009 Pew/JEHT. "Comprehensive Assessment of the Quality of Voter Registration Lists in the United States: A pilot study proposal" (with Alan Gerber). Amount: $100,000.

2010-2011 National Science Foundation, "Cooperative Congressional Election Study," $360,000

2010-2012 Sloan Foundation, "Precinct-Level U. S. Election Data," $240,000.

2012-2014 National Science Foundation, "Cooperative Congressional Election Study, 2010-2012 Panel Study" $425,000

2012-2014 National Science Foundation, "2012 Cooperative Congressional Election Study," $475,000

Professional Boards

Editor, Cambridge University Press Book Series, Political Economy of Institutions and Decisions, 2006-present

Member, Board of the Reuters International School of Journalism, Oxford University, 2007 to present.

Member, Academic Advisory Board, Electoral Integrity Project, 2012 to present.

Contributing Editor, *Boston Review*, The State of the Nation.

Member, Board of Overseers, American National Election Studies, 1999 - 2013.

Associate Editor, Public Opinion Quarterly, 2012 to 2013.

Editorial Board of American Journal of Political Science, 2005 to present.
Editorial Board of Legislative Studies Quarterly, 2005 to present.
Editorial Board of Public Opinion Quarterly, 2006 to present.
Editorial Board of the Election Law Journal, 2002 to present.

Editorial Board of the Harvard International Journal of Press/Politics, 1996 to 2008.
Editorial Board of Business and Politics, 2002 to Present.
Scientific Advisory Board, Polimetrix, 2004 to 2006.

Special Projects and Task Forces

Principal Investigator, Cooperative Congressional Election Study, 2005 – present.

CBS News Election Decision Desk, 2006-present

Co-Director, Caltech/MIT Voting Technology Project, 2000-2004.

Co-Organizer, MIT Seminar for Senior Congressional and Executive Staff, 1996-2007.

MIT Energy Innovation Study, 2009-2010.
MIT Energy Initiative, Steering Council, 2007-2008
MIT Coal Study, 2004-2006.
MIT Energy Research Council, 2005-2006.
MIT Nuclear Study, 2002-2004.
Harvard University Center on the Environment, Council, 2009-present